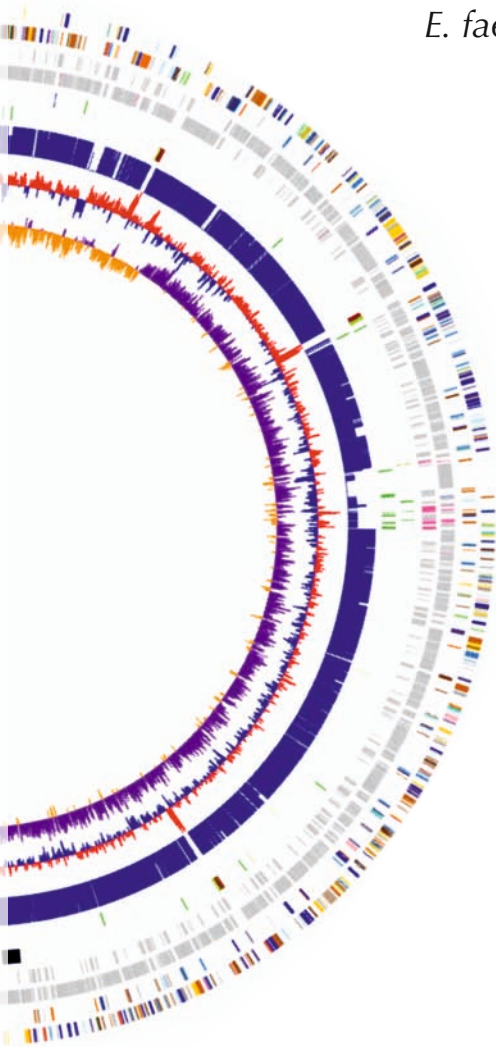


Genomsequenz des probiotischen *Enterococcus faecalis* Symbioflor 1 (DSM 16431) und vergleichende Genomanalyse mit den Stämmen *E. faecalis* V583, *E. faecalis* OG1RF und *E. faecalis* 62

Moritz Fritzenwanker



INAUGURALDISSERTATION

zur Erlangung des Grades eines **Doktors der Medizin**
des Fachbereichs Medizin der Justus-Liebig-Universität Gießen



édition scientifique
VVB LAUFFERSWEILER VERLAG

Das Werk ist in allen seinen Teilen urheberrechtlich geschützt.

Jede Verwertung ist ohne schriftliche Zustimmung des Autors oder des Verlages unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung in und Verarbeitung durch elektronische Systeme.

1. Auflage 2013

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Author or the Publishers.

1st Edition 2013

© 2013 by VVB LAUFERSWEILER VERLAG, Giessen

Printed in Germany

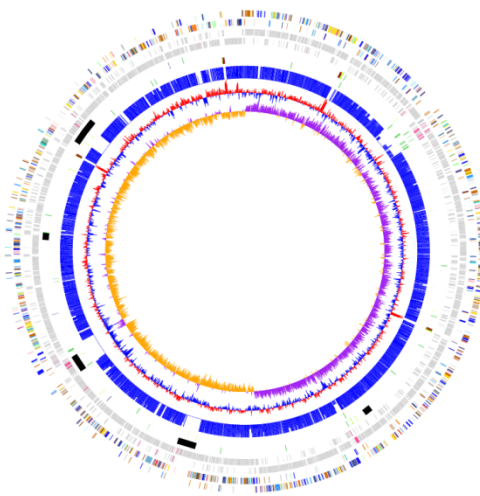


édition scientifique
VVB LAUFERSWEILER VERLAG

STAUFENBERGRING 15, D-35396 GIESSEN
Tel: 0641-5599888 Fax: 0641-5599890
email: redaktion@doktorverlag.de

www.doktorverlag.de

**Genomsequenz des probiotischen
Enterococcus faecalis Symbioflor
1 (DSM 16431) und vergleichende
Genomanalyse mit den Stämmen
E. faecalis V583, *E. faecalis* OG1RF
und *E. faecalis* 62**



Inauguraldissertation

zur Erlangung des Grades eines Doktors der Medizin
des Fachbereiches Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von

Moritz Fritzenwanker

aus Lübeck

Gießen 2012

Aus dem Institut für Medizinische Mikrobiologie

unter Leitung von Prof. T. Chakraborty

Universitätsklinikum Gießen und Marburg GmbH

Standort Gießen

Gutachter: Prof. E. Domann

Gutachter: Prof. Dr. K. T. Preissner

Tag der Disputation: 14.01.2013

I. Inhalt

1	Einleitung	1
1.1	Bakterien – Schädlinge und Nützlinge	1
1.2	Die Gattung <i>Enterococcus</i>	5
1.2.1	Historie der Gattung <i>Enterococcus</i>	5
1.2.2	Die Eigenschaften der Gattung <i>Enterococcus</i>	6
1.2.3	Infektionen durch Enterokokken	7
1.2.4	Potentielle Pathogenitätsfaktoren des <i>E. faecalis</i>	11
1.3	Zielsetzung dieser Arbeit	13
2	Material	15
2.1	Stämme	15
2.2	Kulturmedien	15
2.3	Chemikalien und Materialien	15
2.4	Geräte	16
2.5	Software	17
2.5.1	Lokal ausführbare Software	17
2.5.2	Webbasierte Software & Ressourcen	17
3	Methoden	19
3.1	Sequenzierstrategie für <i>E. faecalis</i> Symbioflor 1	19
3.2	Isolierung chromosomaler DNA gemäß Pitcher	23
3.2.1	Materialien	23
3.2.2	Protokoll	23
3.3	Polymerase-Kettenreaktion (PCR)	25
3.3.1	Expand High Fidelity PCR (Roche)	26
3.3.2	Manual PCR Extender System (5Prime)	27
3.3.3	Phusion Flash High Fidelity PCR System (Finnzymes)	28
3.3.4	Aufreinigung von PCR-Produkten mittels QIAquick PCR Purification Kit ...	28
3.3.5	Sequenzierung der PCR-Produkte	28
3.4	Gelelektrophorese von DNA-Fragmenten	29
3.4.1	Herkömmliche Agarose-Gelelektrophorese	29
3.4.2	Pulsfeld-Gelelektrophorese (PFGE)	30
3.5	Erstellung einer Fosmidbibliothek von <i>E. faecalis</i> Symbioflor 1	31

3.5.1	Herstellung der Fosmid-Bibliothek	31
3.5.2	Fosmid-Restriktionsverdau	38
3.5.3	Screening der Fosmidbibliothek.....	38
3.5.4	Glycerinkulturen	41
3.5.5	Isolierung der Fosmide.....	41
3.6	Bioinformatische Verarbeitung	43
3.6.1	Newbler Assembler (Genome Sequencer FLX Software, Roche).....	44
3.6.2	Mapping der Contigs mit MAUVE	45
3.6.3	Primerwalking mit minimap	46
3.6.4	Assemblierung von PCR-Reads	47
3.6.5	Primerwalking mittels Seqman.....	50
3.6.6	Primerwalking mittels Primer3Plus.....	50
3.6.7	Annotation des Genoms.....	51
4	Ergebnisse	57
4.1	Assemblierung des Genoms von <i>E. faecalis</i> Symbioflor 1.....	57
4.1.1	454-Sequenzierung.....	57
4.1.2	Assemblierung mit Newbler	57
4.1.3	Mapping mit MAUVE	57
4.1.4	Primerwalking mit minimap	57
4.1.5	Primerwalking mit Seqman und Primer3Plus	58
4.1.6	Lückenschluss mittels PCR.....	58
4.1.7	Sequenzierung der rRNA-Operons	58
4.1.8	Fosmid-Bibliothek von <i>E. faecalis</i> Symbioflor 1	59
4.1.9	Finaler Stand der Assemblierung.....	62
4.2	Allgemeine Charakteristika des Genoms von <i>E. faecalis</i> Symbioflor 1.....	63
4.3	Multi-Locus-Sequence-Typing	66
4.4	Phylogenetische Untersuchung	67
4.5	Sequenz-Vergleich der Genome	68
4.5.1	Synteny der Codierenden Sequenzen (CDS)	70
4.5.2	Statistische Analyse der genetischen Homologie	70
4.5.3	COG-Klassifikation der Gene.....	73
4.5.4	Spezifische Gene des <i>Enterococcus faecalis</i> Symbioflor 1.....	74
4.5.5	Putative Pathogenitätsfaktoren	79
4.6	Die verbleibenden Lücken	88

4.6.1	Übersicht über Gap 6	88
4.6.2	Übersicht über Gap 20	92
5	Diskussion	95
5.1	Problemstellung	95
5.2	Lückenschluss	98
5.3	Fazit für die Strategie zum Lückenschluss prokaryontischer Genome	103
5.4	Putative Pathogenitätsfaktoren	106
5.5	Überleben der Magen-Darm-Passage	112
5.6	Putative probiotische Eigenschaften	114
5.7	Ausblick	117
6	Zusammenfassung	119
7	Summary	121
8	Eidesstattliche Erklärung	123
9	Lebenslauf	125
10	Danksagung	127
11	Literaturverzeichnis	129
12	Publikationen	141
13	Anhang	143
13.1	Primer der ersten Runde Lückenschluss-PCRs (minimap)	143
13.2	Primer der zweiten Phase Lückenschluss-PCRs (minimap, Seqman, Primer3+)	144
13.3	Primer für das Fosmid-Screening	146
13.4	Primer für die rRNA-Kassetten	146
13.5	Enterokokken des DMSZ	148

II. Glossar

Die Fachtermini der Genomsequenzierung sind im alltäglichen Gebrauch oft von englisch-sprachiger Literatur geprägt. Für diese Arbeit wurden die meisten der im Haus üblichen englischen Begriffe übernommen, anstatt deutsche ungebräuchliche Entsprechungen dafür zu finden. Folgend einige kurze Erläuterungen:

Sequenzierung: Bedeutet in diesem Fall, die Abfolge von Nukleotiden (DNA-Basen) in einem DNA-Strang zu bestimmen. Dies geschieht mittels verschiedener Techniken, in der Regel wird die DNA zunächst isoliert, gereinigt und chemisch bzw. biochemisch behandelt und dann von einer Sequenzier-Maschine bearbeitet, welche die zugrunde liegende Abfolge von Basen ermittelt.

Read: Eine Sequenz (Abfolge) von DNA-Basen, die von einem Gerät zur DNA-Sequenzierung ausgegeben wird. In der Regel wird eine Datei geliefert, die neben der reinen Sequenz die zugrundeliegenden Rohdaten, z.B. ein Chromatogramm, enthält. Ein Read wird in der Regel von einem Computerprogramm interpretiert, zusätzlich kann ein menschlicher Benutzer aber anhand des Chromatogramms Korrekturen an der Basensequenz vornehmen.

Basecall: Die generierten Rohdaten einer DNA-Sequenzier-Maschine werden von einem Programm interpretiert und es wird daraus eine Nukleotid-Sequenz generiert, die in Form einer Abfolge von Buchstaben als sogenannter Read ausgegeben wird. Der Vorgang, einer bestimmten Position innerhalb der Rohdaten eine Base (bzw. ein Nukleotid) zuzuweisen, wird als basecalling bezeichnet. Die Nukleotide innerhalb des ausgegebenen Reads sind hierbei nur bedingt korrekt, sie sind das Resultat von Berechnungen auf Basis der zugrunde liegenden Rohdaten. Es wird jedoch für jede Base ein Qualitätswert (*Q-value*) angegeben, der eine statistische Aussage über die Richtigkeit der betreffenden Base darstellt.

Trimming: Bei Verwendung von Sequenziertechniken wie Sanger oder 454 sind insbesondere der Anfang und das Ende der Reads stark fehlerhaft, d.h. die statistische Richtigkeit des basecallings wird deutlich schlechter. Daher werden die Reads in der Regel für die weitere Verwendung getrimmt, das heißt die schlechten Sequenzen werden abgeschnitten. Dies kann automatisch geschehen indem verschiedene Kriterien für die Qualität und Länge der zu entfernenden Sequenzen festgelegt werden, oder manuell, indem der Benutzer sich die zugrunde liegenden Chromatogramme anschaut. Im Falle einer Hochdurchsatz-Technologie wie 454 wird wegen der riesigen Anzahl an Reads automatisch getrimmt. Im Falle der

sequenzierten PCR-Produkte wurde bei diesem Projekt in der Regel manuell getrimmt.

Assembly (auch assemblieren): Mit der heute verfügbaren Technik ist es nicht möglich, die DNA-Sequenz eines kompletten bakteriellen Genoms in einem Stück zu bestimmen. Daher werden viele kleine Stücke des Genoms einzeln sequenziert und dann anhand von Überlappungen zu einem Ganzen zusammengefügt. Aufgrund der großen Datenmenge werden hierfür Computerprogramme verwendet, so genannte Assembler. Der Prozess bzw. das Produkt selbst wird als Assembly bezeichnet.

Contig: Bei einem Assembly werden viele einzelne Reads, deren Sequenzen teilweise identisch sind, überlappend aneinandergesetzt. Eine solche Sequenz, die aus mehreren Reads gebildet wurde, wird als Contig bezeichnet. In der Regel entstehen bei einem Assembly der Reads aus einer Hochdurchsatz-Sequenzierung von einem unbekannten bakteriellen Genom mehrere Contigs, die zunächst nicht miteinander verbunden werden können, das heißt, es verbleiben noch Lücken (Gaps).

Consensus: Aufgrund der Fehler, die an unterschiedlichen Stellen innerhalb des ganzen Sequenziervorgangs entstehen können, bilden die meisten Reads die Realität nicht absolut korrekt ab. Dementsprechend sind auch die Überlappungen zwischen zwei Reads nicht unbedingt zu 100% identisch. In der Tat können verschiedene Kriterien für den Assembler festgelegt werden, anhand welcher dieser entscheidet, ob zwei Reads überlappen oder nicht. Die wichtigsten Parameter sind in der Regel die Mindestlänge und die mindestens erforderliche prozentuale Identität der zu überlappenden Fragmente. Hat das Programm nun verschiedenste Reads übereinander gelegt und zusammengefügt, muss es für jede Stelle des Contigs eine Base angeben. Da die Reads wie erwähnt Fehler an den unterschiedlichsten Stellen haben und unter Umständen auch fehlerhaft zusammengefügt sein können (zu niedrige Grenzwerte für Identität), hat das Programm an vielen Positionen des Contigs unterschiedliche Informationen, welche Base an diese Stelle gehört. Das Programm weist dann für diese Position eine sogenannte Consensus-Base zu, in der Regel eine Mehrheitsentscheidung bezüglich der hier zugrunde liegenden Reads. Die Kriterien, welche vom Programm für die Entscheidungsfindung genutzt werden, können hierbei auch, je nach Programm, teilweise gesteuert werden.

Die resultierende gesamte Basen-Sequenz eines Contigs bildet so die Summe der Basen-Informationen vieler Reads und wird auch im gesamten als Consensus-Sequenz bezeichnet.

Well: Vertiefung in einer Mikrotiter-Platte.

III. Abkürzungsverzeichnis

*.ab1	Read-Datei im Format von Applied Biosystem
*.embl	Datei des EMBL-Formates
*.fna	Datei im FASTA-Format
*.gbk	Datei im Genbank-Format
*.scf	Read-Datei im Format nach Staden, Dear 1992
*.sff	Datei des Standard Flowgram Formats
*.txt	Textdatei
AS	Aminosäure
B&D	Lösung für die Isolierung von Fosmiden, siehe Protokoll
bp	Basenpaar
CDS	codierende Sequenz
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Desoxyribonukleinsäure (deoxyribonucleic acid)
DNAse	DNA abbauendes Enzym
dNTP	Desoxyribonukleosidtriphosphat
EDTA	Ethylendiamintetraessigsäure
EHEC	Entero-hämorrhagischer E.coli
EMBL	European Molecular Biology Laboratory
ESBL	Extended-Spectrum-beta-Lactamase
h	Stunde
Hz	Hertz, Schläge pro Sekunde
K	Kalium
KBE	Koloniebildende Einheiten
kbp	Kilobasenpaar
LB	Lysogeny Broth, auch "Luria-Bertani" Medium (bakterielles Nährmedium)
Mg	Magnesium
min	Minute
MLST	Multi-Locus-Sequence-Typing
MRSA	Methicillin-resistenter Staphylococcus aureus

Na	Natrium
NaCl	Natriumchlorid
NaOH	Natronlauge
NCBI	National Center for Biotechnology Information; US-Institution
NGS	Next-Generation-Sequencing
nt	Nukleotid
OD	Optische Dichte
ORF	offenes Leseraster (open reading frame)
oriC	origin of Replication (Replikationsursprung)
PCR	Polymerase-Ketten-Reaktion (polymerase chain reaction)
PFGE	Pulsfeld-Gelelektrophorese
Q40	Qualitätswert für die Richtigkeit einer Base. Entspricht maximal einem Fehler bei 10.000 Basen
RNA	Ribonukleinsäure (ribonucleic acid)
RNAse	RNA abbauendes Enzym
rpm	Umdrehungen pro Minute (rotations per minute)
SDS	Sodiumdodecylsulfat, auch Natriumdodecylsulfat
sec	Sekunde
SNP	Single-Nucleotid-Polymorphism
TAE	Tris-Acetat-EDTA-Puffer
Taq	<i>Thermus aquaticus</i>
TBE	Tris-Borat-EDTA-Puffer
TE	Tris-EDTA-Puffer
Tgo	<i>Thermococcus gorgonarius</i>
TNE	Tris-NaCl-EDTA-Puffer
Tris	Tris(hydroxymethyl)-aminomethan
TSB	Tryptic Soil Broth (bakterielles Nährmedium)
VRE	Vancomycin-resistente Enterokokken
WHO	World Health Organization

IV. Abbildungsverzeichnis

Abbildung 1: Elektronenmikroskopische Aufnahme von <i>E. faecalis</i> Symbioflor 1.....	6
Abbildung 2: Strategie zur Zielsetzung	22
Abbildung 3: Vektor pcc1Fos.....	32
Abbildung 4: Herstellung der Fosmid-Bibliothek	33
Abbildung 5: Hamilton-Spritze	34
Abbildung 6: Größenselektion der DNA.....	35
Abbildung 7: Fosmid-Klone nach erfolgreicher Infizierung.....	36
Abbildung 8: Lückenschluss-PCRs.....	58
Abbildung 9: Die 4 rRNA-Operons des <i>E. faecalis</i> Symbioflor 1 in MAUVE	59
Abbildung 10: Fosmid-Genbank-Platte.....	59
Abbildung 11: Fosmid-Restriktionsverdau mit BamH1	60
Abbildung 12: Fosmid-Screening-PCRs	61
Abbildung 13: Übersicht über das Genom von <i>E. faecalis</i> Symbioflor 1	64
Abbildung 14: Phylogenetischer Baum anhand von MLST.....	66
Abbildung 15: Phylogenetischer Baum der 4 <i>E. faecalis</i> -Stämme.....	67
Abbildung 16: Homologie-Vergleich der <i>E. faecalis</i> -Stämme mit MAUVE.....	68
Abbildung 17: Pangenom der proteinkodierenden Sequenzen der 4 Enterokokken- Stämme	70
Abbildung 18: Euler-Diagramm zur Visualisierung der Homologie	72
Abbildung 19: Klassifikation der Gene nach der COG-Klassifikation.....	73
Abbildung 20: Lactose/Cellobiose-spezifisches PTS-System.....	75
Abbildung 21: Potentielles Insertionselement mit Restriktionsenzym und Genen für Zucker-Metabolismus	76
Abbildung 22: Region um den Multidrug-ABC-transporter.....	77
Abbildung 23: Gene im Bereich der Aggregation Substance	79
Abbildung 24: Gene im Bereich des Collagen Adhesion Protein	79
Abbildung 25: Gene im Bereich der Gelatinase.....	80
Abbildung 26: Gene im Bereich des Enterococcal Surface Proteins	81
Abbildung 27: Pathogenitätsinsel 1	81

Abbildung 28: Pathogenitätsinsel 2	82
Abbildung 29: <i>vanG</i> -ähnlicher Locus.....	83
Abbildung 30: Katalase - fragmentiertes Gen	85
Abbildung 31: Thioredoxin-Reduktase - fragmentiertes Gen	85
Abbildung 32: Das <i>cps</i> -Operon.....	86
Abbildung 33: Das <i>epa</i> -Operon	86
Abbildung 34: CRISPR-Loci	87
Abbildung 35: PCRs und Chromatogramme von Gap 6	89
Abbildung 36: Bestimmung der Größe von Gap 6	90
Abbildung 37: Übersicht über Gap 6.....	91
Abbildung 38: Übersicht über Gap 20.....	93
Abbildung 39: Alternative Ausrichtung der Contigs.....	100
Abbildung 40: Das Problem, repeats zu assemblieren	105

1 Einleitung

1.1 Bakterien – Schädlinge und Nützlinge

Die Vorstellung von Bakterien innerhalb der Humanmedizin als schädigendes Agens ist wohl landläufig die bekannteste. Und in der Tat machen Bakterien ihrem Ruf als mikroskopisch kleine „Killer“ auch heute noch alle Ehre. Die WHO gab 2008 eine Studie heraus, in welcher statistische Untersuchungen von Krankheits- und Todesursachen weltweit veröffentlicht wurden (WHO 2008). Demnach sind Infektionskrankheiten die zweithäufigste Todesursache nach den kardiovaskulären Krankheiten, gefolgt von Krebs an dritter Stelle. Sie sind für 15,6% der weiblichen und für 16,7% der männlichen Todesfälle verantwortlich. Eine genauere Auftrennung der Ursachen zeigt Infektionen der unteren Atemwege an dritter Stelle mit 7,1% der Todesfälle, Durchfallerkrankungen mit 3,7% an fünfter Stelle, und die Tuberkulose mit 2,5% an siebter Stelle als infektiös bedingte Ursachen. Es gibt wesentliche Unterschiede zwischen den einzelnen Ländern: In den Ländern mit niedrigem Einkommen machen die Atemwegsinfektionen mit 11,2% sogar den ersten Platz, die Durchfallerkrankungen mit 6,9% den dritten, und die Tuberkulose mit 3,5% den siebten Platz – sie stehen damit sogar vor der Malaria mit 3,3%. In den Ländern mit hohem Einkommen jedoch kommen die Atemwegsinfektionen mit 3,8% noch auf den vierten Platz, Tuberkulose und Durchfallerkrankungen sind nicht mehr unter den zehn häufigsten Ursachen. In ärmeren Ländern sind immer noch vor allem Kinder von Bakterien bedroht: Infektionen der Atemwege, vor allem Pneumonien, sind für 17% der Todesfälle weltweit der Kinder unter 5 Jahren verantwortlich. *Streptococcus pneumoniae* und *Hämophilus influenzae* sind hierbei die häufigsten Erreger, beides Bakterien. An zweiter Stelle mit 16% stehen die Durchfallerkrankungen. Neben dem Rotavirus als häufigstem Erreger kommen hier *E. coli*, *Shigella*, *Campylobacter* und *Salmonella* als bakterielle Erreger in Betracht, ebenso gibt es immer wieder Epidemien, die von *Vibrio cholerae* ausgelöst werden. Vor allem in Situationen, in denen Menschen aufgrund von Naturkatastrophen oder humanitären Krisen in engen Unterkünften mit schlechter hygienischer Versorgung, insbesondere unsauberem

Trinkwasser oder schlechter sanitärer Situation, leben müssen, kommt es zu vielen durchfallbedingten Todesfällen. Malaria, zum Vergleich, ist für nur 7% der Todesfälle bei Kindern unter 5 Jahren verantwortlich (Wardlaw et al. 2010).

In den reicheren Ländern haben verbesserte Hygiene-Maßnahmen und die breite Verfügbarkeit von Antibiotika zu einer drastischen Reduktion der bakteriell bedingten Sterblichkeit geführt. Nichtsdestotrotz kommt es auch hierzulande immer noch zu bakteriell bedingten Todesfällen. Aktuell bleibt der Ausbruch von EHEC-Bakterien (Enterohämorrhagischer *E. coli*) in Erinnerung: Im Jahr 2011 kam es zu einem Ausbruch, bei dem insgesamt 3842 Erkrankungen durch den Stamm O104:H4 gemeldet wurden, davon beinhalteten 855 Fälle die Entwicklung eines hämolytisch-urämischen Syndroms. Zusätzlich wurden 719 EHEC-Fälle gemeldet, die nicht dem Ausbruch zugeordnet wurden. Das mediane Erkrankungsalter bezogen auf den Ausbruch lag bei 46 Jahren. Zuvor galt EHEC vornehmlich als pädiatrisches Problem, in den Jahren 2006-2010 wurden im gleichen Zeitraum im Median 218 Fälle von EHEC-Gastroenteritis und 13 Fälle von hämolytisch-urämischem Syndrom berichtet. Während des Ausbruchs starben 18 Personen (0,6%) an EHEC-bedingter Gastroenteritis und 35 Personen (4,1%) an durch EHEC bedingtem hämolytisch-urämischem Syndrom – Zahlen, die angesichts der erwähnten Kindersterblichkeit klein erscheinen, Bakterien als tödliche Krankheitserreger jedoch erneut ins Licht des öffentlichen Interesses und der Medien gebracht haben (Robert-Koch-Institut 2011).

Nicht nur in solchen medienwirksamen Ausbrüchen sind bakterielle Infektionen auch hierzulande noch von Bedeutung. Als Beispiel sei die ambulant erworbene Pneumonie genannt, eine vornehmlich bakteriell bedingte Erkrankung. Nach Schätzungen sind in Deutschland jährlich 400.000 bis 600.000 Menschen von dieser Erkrankung betroffen. Die Letalität wird mit 0,6% bei ambulanten Fällen und mit 13,7 bis 14,4% bei stationär behandelten Patienten angegeben. Die Inzidenz steigt mit zunehmendem Alter. (Höffken et al. 2009)

Der Anteil der Bakterien, der Krankheiten verursacht, macht jedoch nur einen Bruchteil der vorhandenen Bakterien aus. Mensch und Tier sind physiologischerweise von einer Vielzahl von Bakterien besiedelt, ohne dass von einer primär pathogenen Wirkung auszugehen ist. Die Gesamtzahl der den Körper besiedelnden Mikroorganismen wird als Mikroflora bezeichnet, international wird hierfür aktuell der Begriff Mikrobiota verwendet.

Der Mensch besteht aus etwa 10^{13} eigenen Zellen und trägt etwa 10^{14} bakterielle Zellen mit sich, damit kommen auf jede körpereigene Zelle rein rechnerisch 10 bakterielle. Insbesondere der Verdauungstrakt ist von einer Vielzahl an verschiedensten Bakterien besiedelt. So tragen Magen und Duodenum 10^1 - 10^3 KBE/ml, Jejunum und Ileum 10^4 - 10^8 KBE/ml und das Kolon 10^{10} - 10^{12} KBE/g (Sonnenborn & Greinwald 1991).

Diese Mikrobiota ist nicht nur einfach eine unspezifische Sammlung an Bakterienarten, sondern vielmehr eine im Laufe der Ko-Evolution von Tieren und Bakterien entstandene Zweckgemeinschaft mit spezifischer Zusammensetzung und Aufgabenverteilung. So haben Lebewesen in unterschiedlichen Lebensräumen mit unterschiedlichem Nahrungsangebot unterschiedliche Mikrobiota entwickelt, die auf die jeweiligen Bedürfnisse angepasst sind. (Ley et al. 2009)

Die Mikrobiota eines Menschen erfüllt verschiedene wichtige Funktionen: Sie bildet eine Art natürliche Konkurrenz gegenüber potentiell schädlichen Keimen, sie beeinflusst die Entwicklung des Immunsystems, sie verhindert, dass zu viele Mikroorganismen in das Lymphsystem gelangen, sie baut Ballaststoffe aus der Nahrung und Abfallprodukte des Körpers ab, sie produziert kurzkettige Fettsäuren, die von Kolon-Zellen zur Ernährung verwendet werden, zudem trägt sie zur Verbesserung der Darmschleimhautdurchblutung, Darmmotilität und Resorption von Wasser und Elektrolyten bei, sie entgiftet endogene und exogene toxische oder karzinogene Stoffe, sie produziert Vitamine und sie regelt den pH-Wert des Darm-Milieus. Die Bakterien der Mikroflora kommunizieren sowohl mit dem Darmepithel als auch untereinander – so wird zum Beispiel die Kolonisationsdichte über Quorum Sensing beeinflusst.

Ursprünglich wurde vermutet, dass durch den Abbau von Eiweißen Fäulnisprodukte im Dickdarm entstehen, die den Menschen langfristig vergiften würden (H. Senator 1868, nach Schulze et al. 2008).

Im Jahr 1908 wurde von Ilja Metschnikoff „The Prolongation of life“ publiziert, in welchem er vorschlug, *Lactobacilli* prophylaktisch einzusetzen, um diesem schädlichen Einfluss entgegenzuwirken. Obwohl spätere Untersuchungen dieses Krankheitsmodell widerlegten, kann diese Behandlungsweise als geschichtliche Grundlage für die Probiotika interpretiert werden, denn die Grundidee ist, durch orale Einnahme von lebenden Bakterien eine Verbesserung der Gesundheit zu erreichen.

Dem entspricht in etwa die heutige Definition der Food and Agriculture Organization (FAO) und der WHO von 2001: „Lebende Mikroorganismen, die, wenn in ausreichender Menge verabreicht, dem Wirtsorganismus einen gesundheitlichen Nutzen bringen“ (Schulze, 2008).

Aktuell wirbt zum Beispiel die Firma Danone mit der gesundheitsfördernden Wirkung von *Lactobacillus casei* *defensis*, einem probiotischen Stamm, der in den Trinkjogurts der Marke „Actimel“ enthalten ist (www.actimel.de) – ein Stamm, der wissenschaftlich eigentlich den Namen *L. casei* DN-114001 trägt.

Wenn man gemeinhin von Probiotika spricht, ist zu unterscheiden, ob es sich um ein Arzneimittel – mit nachgewiesener Wirkung – handelt, oder um einen Lebensmittelzusatzstoff oder ein Nahrungsergänzungsmittel – nicht unbedingt mit nachgewiesener Wirkung.

Da Lebensmittel in Deutschland nicht zur Behandlung von Krankheiten beworben werden dürfen, ohne dass ein wissenschaftlicher Nachweis hinsichtlich Wirksamkeit erbracht wurde, wird zur Vermarktung oftmals auf den Begriff „Wellness“ zurückgegriffen. Damit kann das Produkt vermarktet werden, ohne eine definierte Wirksamkeit bezüglich der Prophylaxe oder Behandlung von Krankheiten nachweisen zu müssen. Der Begriff Probiotikum an sich verspricht somit keine gesicherte Wirkung, sofern es sich nicht um ein Arzneiprodukt handelt (Schulze et al. 2008).

Vergleichsweise gut abgesicherte Wirkungen beim klinischen Einsatz von Probiotika gibt es bei akuten und bei durch Antibiotika ausgelösten Gastroenteritiden sowie bei allergischen Reaktionen, insbesondere der atopischen Dermatitis. Zudem gibt es erfolgsversprechende Ergebnisse bei kindlichen Atemwegsinfektionen, Karies, Infektionen der Nasennebenhöhlen, und bei wiederkehrenden Infektionen mit *Clostridium difficile* und anderen entzündlichen Darmkrankheiten. Ebenso gibt es Hinweise auf einen möglichen Nutzen bei Rheumatoider Arthritis und beim Reizdarmsyndrom (Goldin & Gorbach 2008).

Die meisten probiotischen Medikamente verwenden Milchsäurebakterien wie *Lactobacilli* (Goldin & Gorbach 2008; Schulze et al. 2008). Die Gattung *Lactobacillus* wird im Allgemeinen als apathogen angesehen, aber auch sie wird, wenn auch eher selten, als Erreger bei Harnwegsinfekten oder Bakteriämien gefunden. Dies mag

jedoch unter anderem daran liegen, dass sie dem kulturellen Standard-Nachweis nicht immer zuverlässig zugänglich sind (Imirzalioglu et al. 2008).

Während es beim *Lactobacillus* eher die Ausnahme als die Regel ist, Infektionen auszulösen, verhält sich der *Enterococcus* wie der römische Gott Janus mit seinen zwei Gesichtern: Einerseits ist er aus der „normalen“ Mikrobiota des Darms kaum wegzudenken, wird gar als gesundheitsförderliches Probiotikum verkauft, andererseits tritt er regelmäßig als Erreger von Infektionen auf.

1.2 Die Gattung *Enterococcus*

1.2.1 Historie der Gattung *Enterococcus*

Die erste Erwähnung des Begriffs „entérocoque“ findet sich bei Thiercelin in einer französischen Publikation aus dem Jahre 1899, er beschrieb grampositive Diplokokken intestinalen Ursprungs. Im Jahre 1906 wurden von Andrewes und Horder bei einem Patienten mit Endokarditis Bakterien isoliert, die sie „*Streptococcus faecalis*“ nannten, da sie charakteristisch für den menschlichen Darm seien. Einige Jahre später wurden von Orla-Jensen Stämme mit den Namen *S. glycerinaceus* und *S. faecium* beschrieben. Eine Zeit lang wurde der Name *Enterococcus* teilweise synonym zu *S. faecalis* und teilweise generell für aus den Fäzes isolierte Streptokokken verwendet. Sherman schlug 1937 eine Einteilung der *Streptokokken* in 4 Gruppen vor: *pyogene*, *viridans*, *lactic* und *enterococcus*. Die Einteilung erfasste Organismen, die zum Großteil bei 10 und 45°C, in 6,5% NaCl, und bei einem pH-Wert von 9,6 Wachstum zeigten, sie überlebten 60°C für 30 min, außerdem besaßen sie die Fähigkeit Äskulin zu hydrolysieren. Zudem zeigte sich, dass die Enterokokken der damaligen Definition mit den Antisera der Lancefield-Gruppe D reagierten. Zu den *Enterococci* zählten die Spezies *S. faecalis*, *S. zymogenes*, *S. liquefaciens* und *S. durans*. In den 1940er und 1950er Jahren zeigten Studien, dass der von Sherman als identisch zum *S. faecalis* betrachtete *S. faecium* aufgrund von biochemischen Eigenschaften als eine eigene Spezies zu interpretieren sei. Im Verlauf der Jahre wurden weitere Organismen entdeckt, die zu den *Enterococci* gezählt wurden: *S. faecium* subsp. *mobilis*, *S. faecium* var. *casseliflavus*, *S. faecalis* var. *malodoratus*, *S. avium* sowie *S. gallinarum*. Zudem wurde festgestellt,

dass *S. faecalis*, *S. liquefaciens* und *S. zymogenes* als eine Spezies zu betrachten seien.

Im Jahr 1984 zeigten Schleifer und Kilpper-Bälz durch DNA-DNA und DNA-rRNA-Hybridisierung, dass *S. faecalis* und *S. faecium* so weit entfernt verwandt mit der Gattung *Streptococcus* sind, dass für sie eine eigene Gattung geschaffen werden sollte. Sie führten den Namen *Enterococcus* als Gattung ein. Kurze Zeit später zeigten Collins, Jones und Farrow mit ähnlichen Methoden, dass *S. avium*, *S. casseliflavus*, *S. durans*, *S. faecalis* subsp. *malodoratus* und *S. gallinarum* vom Verwandtschaftsgrad ebenfalls zu jener Gattung zu zählen seien und benannten sie in der Folge um zu *Enterococcus avium*, *E. casseliflavus*, *E. durans*, *E. malodoratus* und *E. gallinarum*. Spätere Studien fügten der Gattung die Spezies *E. hirae*, *E. raffinosus*, *E. solitarius* und *E. pseudoavium* hinzu (B. E. Murray 1990)

Einen Überblick über die verschiedenen Spezies der Gattung *Enterococcus* bietet die Tabelle „Enterokokken vom DSMZ“ im Anhang; sie listet die in der Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH vorhandenen Stämme.

1.2.2 Die Eigenschaften der Gattung *Enterococcus*

Die Mitglieder der *Enterococci* sind grampositive, fakultativ anaerobe Bakterien von ovoider Form. Sie zählen zu den Milchsäurebakterien. Sie ordnen sich einzeln, in Pärchen oder in kurzen Ketten an. Die meisten wachsen in 6,5% NaCl-Lösung und bei einem pH-Wert von 9,6, sie wachsen bei 10°C und bei 45°C, und überstehen 60°C für 30min.

Zur Abgrenzung gegenüber den Staphylokokken dient der Katalase-Test: Er fällt sowohl bei Streptokokken als auch Enterokokken negativ aus, bei Staphylokokken positiv.

Zudem reagieren Enterokokken mit Antisera der Lancefield-Gruppe D, manche auch mit Gruppe Q.

Als metabolische Charakteristika werden genannt: Fähigkeit zur Spaltung von Äskulin in der

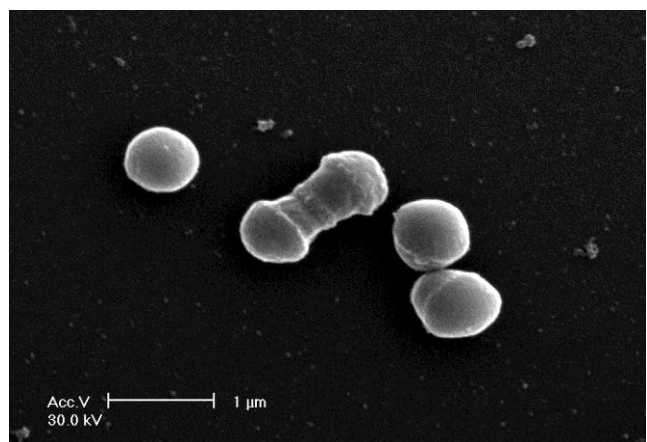


Abbildung 1: Elektronenmikroskopische Aufnahme von *E. faecalis* Symbioflor 1

mit freundlicher Genehmigung von Dr. Martin Hardt, BFS Imaging Unit

Gegenwart von Galle, Fähigkeit zur Spaltung von L-Pyrrolidonyl- β -Naphthylamid. Eine Möglichkeit zur Differenzierung besteht darin, Wachstum in Medium mit L-Pyrrolidonyl- β -Naphthylamid nachzuweisen; diese Fähigkeit besitzen sowohl Enterokokken als auch A-Streptokokken. Über den serologischen Test mit Lancefield-Antiserum können die beiden unterschieden werden (B. E. Murray 1990).

Habitat: *Enterococci* gelten als klassische Stuhlkeime. Sie können in Fäzes der meisten erwachsenen Menschen gefunden werden. Die häufigsten Vertreter sind *E. faecalis* und *E. faecium*. Wie aus der Übersicht der *Enterokokken* des DMSZ hervorgeht, können die verschiedenen Spezies auch aus den Fäzes von Tieren stammen oder in der Umwelt, zum Beispiel in Gewässern, gefunden werden. Gelegentlich können *Enterokokken* auch in der Vaginal- und Mundflora nachgewiesen werden (B. E. Murray 1990).

1.2.3 Infektionen durch Enterokokken

1.2.3.1 Harnwegsinfektionen

Harnwegsinfektionen sind meistens endogener Natur. Bakterien aus dem Enddarm gelangen über die Harnröhre in die Blase und lösen eine Entzündung aus. Harnwegsinfekte kommen deshalb vor allem bei Frauen häufig vor, weil die weibliche Anatomie diesen Übertritt leichter zulässt.

Geschlechtsverkehr, Blasenkatheter oder iatrogene Manipulationen erhöhen das Risiko einer Infektion.

Die häufigste Variante eines Harnwegsinfekts ist die Zystitis, hier bleibt die Infektion auf die Blase beschränkt (unterer Harnwegsinfekt). Typische Symptome sind brennende Schmerzen beim Wasserlassen (Algurie), häufiger Harndrang mit oft nur kleinen Portionen (Pollakisurie), allgemein erschwerte Miktion (Dysurie). Bei einer akuten Zystitis können krampfartige Schmerzen im Unterbauch hinzukommen. Auch Allgemeinsymptome wie Schwäche und Übelkeit können auftreten.

In den meisten unkomplizierten Fällen heilt die Zystitis innerhalb weniger Tage spontan ab. Sie ist weniger ein bedrohliches Problem als ein häufiges: Unter jungen, sexuell aktiven, nicht schwangeren Frauen kann von einer Inzidenz von etwa 0,7 pro Patientin und Jahr ausgegangen werden. Zudem ist in 1-2% der Schwangerschaften

mit einer Zystitis zur rechnen. Bei ansonsten gesunden Frauen in der Postmenopause kann von 6,7 Harnwegsinfektionen pro 100 Patientinnenjahren ausgegangen werden (Wagenlehner et al, 2010).

Die gefährlichere Variante ist der obere Harnwegsinfekt, die Pyelonephritis, bei der es zu einer Beteiligung des Nierenparenchyms kommt. Symptome sind klassischerweise Flankenschmerzen und Klopfschmerz in den Nierenlagern, zudem Fieber und in der Regel schlechteres Allgemeinbefinden als bei einer Zystitis, allerdings ist dies klinisch nicht immer sicher zu unterscheiden.

Das Nierenparenchym kann hierbei bleibend geschädigt werden mit der Folge einer Einschränkung der Nierenfunktion. Zudem kann es akut zu einer Streuung der Bakterien ins Blut kommen mit Folge einer Urosepsis, einem potentiell tödlichen Krankheitsbild.

In der Schwangerschaft tritt zu 1-2% eine Pyelonephritis auf, was ebenso mit Gefahren für Mutter und Kind einhergehen kann. Im postmenopausalen Alter kann eine Prävalenz von 1,19/1000 pro Jahr als Richtwert gelten (Wagenlehner et al, 2010).

Die Diagnostik beider Erkrankungsformen erfolgt in erster Linie über die Anamnese und die klinische Untersuchung. In der Regel folgt eine Urindiagnostik: Der Urin wird makroskopisch auf Blut und Trübung inspiziert. Eine mikroskopische Untersuchung kann nebst Leukozytenzylindern und Erythrozyten auch die Bakterien selbst zeigen. Üblicherweise wird auch ein Teststreifen zum Nachweis von Nitrit verwendet.

Wichtig ist ebenso die Anzahl der Bakterien im Urin. Hierzu wird ein agarbeschichteter Objektträger in den Urin getaucht und für 24 h bei 37°C inkubiert, dann wird die Besiedelung semi-quantitativ durch Vergleich mit einem Standard ermittelt: Koloniezahlen ab 10^5 /ml gelten als Hinweis auf einen Harnwegsinfekt.

Problematisch in dieser Beziehung ist, dass nicht jede Besiedelung des Harntraktes eine Infektion im pathologischen Sinn bedeutet. Bei vielen Frauen können Bakterien im Urin nachgewiesen werden, ohne dass diese dadurch einen Schaden davontragen. Bei älteren Frauen in Pflegeheimen wurden Raten von 25-50% gefunden. Hinzu kommt, dass der Urin, wenn er von den Patientinnen selbst

gewonnen wird, oft nicht steril gewonnen wird. Hautkeime sorgen so für falsch positive Resultate.

Weder bei jungen noch bei älteren Frauen sollte daher der alleinige Nachweis von Bakterien im Urin Anlass zur Gabe von Antibiotika sein – sofern sie nicht schwanger sind. Bei Schwangeren wird auch die asymptomatische Bakteriurie in der Regel behandelt.

Aufgrund der ätiologischen Vorstellung verwundert es nicht, dass die meisten Harnwegsinfekte von Darmkeimen verursacht werden. Die Zahlen unterscheiden sich in verschiedenen Studien, meist ist jedoch *E.coli* der häufigste Erreger. Dem folgen andere gramnegative Enterbacteriaceae wie *Klebsiella*, *Proteus* und *Enterobacter*, sowie die Gattungen *Pseudomonas*, *Staphylococcus* und *Enterococcus*. Als Orientierung werden die Zahlen aus einem Standardlehrbuch der medizinischen Mikrobiologie zu den Erregern von Harnwegsinfekten genannt (Hof & Dörries, 2005):

- *E. coli* 50-70%
- andere gramnegative Enterobacteriaceae 10-20%
- Enterokokken 10-20%
- *Pseudomonas aeruginosa* 5%
- *Staphylococcus aureus* 5%
- Pilze <5%
- *Enterobius* <1%

Anspruchvoll zu kultivierende Erreger entgehen jedoch dem klassischen kulturellen Nachweis: In einer Studie mit molekularbiologischen Techniken wurde gezeigt, dass in etwas mehr als 22% der Urinproben durch Kultur nicht erfasste Bakterien enthalten waren, wie zum Beispiel der oben erwähnte *Lactobacillus*. Bezüglich der Haupterreger *E. coli* und *E. faecalis* kam die Studie zu vergleichbaren Zahlen (Imirzalioglu et al. 2008).

Es scheint, dass sowohl bei komplizierten Harnwegsinfekten als auch mit steigendem Alter des Patienten vermehrt Enterokokken nachgewiesen werden. Zudem scheinen sie in jüngerer Zeit häufiger nachgewiesen zu werden. (B. E. Murray 1990; Schaberg et al. 1991).

1.2.3.2 Bakteriämie

Der Übertritt von Enterokokken ins Blut tritt gehäuft auf bei Harnwegsinfektionen (Urosepsis), bei Verwendung von Blasen- oder intravaskulären Kathetern, aber auch durch intraabdominelle Wunden, durch Verletzungen der Gallenwege und bei Verbrennungen.

In den letzten Jahrzehnten wurden immer häufiger Enterokokken in Blutkulturen nachgewiesen, insbesondere bei hospitalisierten Patienten (Schaberg et al. 1991). Oft werden neben den Enterokokken zusätzlich andere Bakterien nachgewiesen (24-45%). Die Bakteriämie mit Enterokokken hat eine hohe Mortalität (34-46%). Allerdings ist fraglich, ob dies an den Enterokokken selbst liegt, denn es ist ebenso denkbar, dass erst schwere Grunderkrankungen wie Nierenversagen, maligne Tumoren, Verbrennungen und Alkoholismus den Patienten in einen Zustand bringen, in dem diese Bakterien ins Blut übertreten können (B. E. Murray 1990; Poh et al. 2006).

1.2.3.3 Endokarditis

Im Jahr 1899, in dem der Term „entérocoque“ von Thiercelin eingeführt wurde, wurde bei einem Patienten mit Endokarditis ein Bakterium isoliert, welches von seinen Entdeckern McCallum und Hastings *Micrococcus zymogenes* genannt wurde. Später stellte sich heraus, dass es wahrscheinlich ein *S. faecalis*, also ein *Enterococcus* gewesen war.

Die Endokarditis ist ein ernstes Krankheitsbild: Bakterien aus dem Blutkreislauf setzen sich an den Herzklappen fest, vermehren sich und bilden einen Biofilm. Sie zerstören dabei das Gewebe und verursachen dadurch eine Klappeninsuffizienz. Die Krankheit kann akut oder subakut verlaufen. Symptome können allgemeine Schwäche und Müdigkeit, Fieberschübe und neu aufgetretene Herzgeräusche und Zeichen der Herzinsuffizienz sein. Zusätzlich können Arthralgien, Hautsymptome und Nierenbeteiligung auftreten.

Enterokokken verursachen geschätzte 5 bis 15% der bakteriellen Endokarditiden. Meistens wird *E. faecalis* identifiziert, aber auch *E. faecium*, *E. avium*, *E. casseliflavus*, *E. durans*, *E. gallinarum*, und *E. raffinosus* konnten nachgewiesen werden.

Die Endokarditis durch Enterokokken betrifft meistens die Aorten- und die Mitralklappe. Eine Vorschädigung gilt als Risikofaktor, ist jedoch nicht zwingend erforderlich.

Das Durchschnittsalter wird mit 65 Jahren angegeben, Männer sind häufiger als Frauen betroffen (Murray, 1990).

Eine spanische Multicenter-Studie zum Thema linksseitiger Endokarditiden aus dem Jahr 2009 berichtete, dass von 76 Patienten mit linksseitiger Endokarditis durch Enterokokken 36,8% im Verlauf einen chirurgischen Eingriff an den Herzklappen benötigten. Die Mortalität lag mit 32,9% über den 9,3% bei Infektionen mit *S. viridans* und unter den 48,6% bei *S. aureus* (Martínez-Marcos et al. 2009).

1.2.3.4 Weitere Krankheiten

Neben den genannten Beispielen werden Enterokokken identifiziert als Erreger von Peritonitiden, Wundinfektionen und Katheterbesiedlungen (B. E. Murray 1990).

1.2.4 Potentielle Pathogenitätsfaktoren des *E. faecalis*

Weil der *Enterococcus* im Regelfall in der Mikrobiota von gesunden Menschen vorkommt, kann er kaum als obligat pathogenes Bakterium bezeichnet werden. Die obigen Beispiele machen deutlich, dass er jedoch sowohl häufige als auch schwere Krankheiten auslösen kann. Die Gattung wird daher als fakultativ pathogen eingestuft.

Wenn ein *Enterococcus* seine natürliche Umgebung im Darm verlässt und andere Bereiche wie die Blase infiziert, muss er zunächst Adhäsionsfaktoren besitzen, um an Oberflächen haften zu können. Dann muss er sich einen Weg durch das Gewebe bahnen, sich gegen das Immunsystem und oxidativen Stress erwehren und mit einem geringeren Angebot an Nährstoffen zurechtkommen. Im Rahmen des klinischen Umfelds ist zudem eine Abwehr gegen Antibiotika erforderlich.

Folgende Faktoren geben dem *Enterococcus faecalis* die Fähigkeit pathogener Potenz (Jett et al. 1994; Rich et al. 1999; Domann et al. 2007; Kreft et al. 1992):

- Cytolysin: lysiert grampositive Bakterien und bestimmte eukaryontische Zellen

- Aggregation Substance: ein Oberflächenprotein, das oft auf Pheromon-gesteuerten Plasmiden codiert ist und dazu dient, sich an Zelloberflächen von Epithelien oder auch anderen Bakterien anzuheften und damit zur Aggregation, also Verklumpung führt. Es vermittelt dadurch auch den Austausch von Plasmiden.
- Pheromone: locken *in vitro* Neutrophile an
- Gelatinase: eine Zink-Endopeptidase
- AS-48: ein Bacteriocin mit Wirkung gegen grampositive und gramnegative Bakterien.
- Hyaluronidase: Mucopolysaccharidase
- Lipoteichonsäure (wie die meisten grampositiven Bakterien): stimuliert Zytokin-Produktion in Monozyten; bindet an Aggregation Substance beim Verklumpen
- Collagen Adhesion Protein: Adhäsion an Gewebe
- Enterococcal Surface Protein: Adhäsion an Gewebe
- Widerstandsmechanismen gegen oxidativen Stress
- Kapselbildung
- Biofilmbildung

Im Rahmen des Einsatzes von Antibiotika hat sich *Enterococcus faecalis* im Laufe der Zeit als problematisch erwiesen. Zunächst sind alle Enterokokken natürlicherweise resistent gegen Benzylpenicilline und Cephalosporine. Eingesetzt werden daher Breitbandpenicilline wie Ampicillin, Amoxicillin oder Mezlocillin zusammen mit Aminoglykosiden. Gegebenenfalls kann auch Cotrimoxazol verwendet werden. Gegen diese Antibiotika kann der Keim jedoch Resistenzen entwickeln, sodass unter Umständen auf Vancomycin, ein Reserve-Antibiotikum gegen Kokken, zurückgegriffen werden muss.

Immer häufiger treten jedoch mittlerweile auch Stämme auf, die resistent gegenüber Vancomycin sind, sogenannte VRE (Vancomycin-resistente Enterokokken). Insbesondere im klinischen Bereich, in dem viele Antibiotika verwendet werden, wo außerdem viele Möglichkeiten zum Austausch von Bakterien bestehen, verursachen diese VRE nosokomiale Infektionen, die dann schwer zu beherrschen sind (Hof et al. 2005).

1.2.4.1 Die gute Seite der Enterokokken

Zum einen sind Enterokokken sowieso heimisch in der Darmflora des Menschen. Zum anderen werden sie auch zur Produktion für Nahrung benutzt: Sie tragen zum typischen Geschmack von Käsesorten bei und sind auch in verschiedenen traditionell fermentierten Würstchen, Oliven und Gemüsesorten enthalten (Domann et al. 2007; Jurkovic et al. 2006; Nieto-Arribas et al. 2011). Es gibt sogar Hinweise, dass sie zur Sicherheit von Milchprodukten beitragen, indem sie mit ihrem Bacteriocin andere Bakterien bekämpfen, wie zum Beispiel pathogene Vertreter der Gattung *Listeria* (Chanos & D. R. Williams 2011).

Darüber hinaus finden sie Anwendung in verschiedenen probiotischen Medikamenten, zum Beispiel in Bioflorin (*E. faecium* SF68, Sanova), Cylactin (*E. faecium* SF68, als Futterzusatzmittel, Hoffmann-La Roche, Basel, Schweiz), Fargo 688 (Quest International, Naarden, Niederlande), ECOFLOR (Walthers, Den Haag, Niederlande) und dem in dieser Arbeit untersuchten Symbioflor 1 (*E. faecalis* Symbioflor 1, SymbioPharm, Herborn, Deutschland) (Domann et al., 2007; Schulze et al., 2008).

Für Symbioflor 1 ist bisher nachgewiesen, dass bei der Behandlung einer chronischen Bronchitis ebenso wie bei einer chronischen Sinusitis Rückfälle seltener und später auftreten. Die Wirkung war bis in den Follow-up-Zeitraum von 8 Monaten nachweisbar (Habermann et al. 2002; Habermann et al. 2001). Eine Studie an Zellen aus dem peripheren Blut zeigte, dass der probiotische Keim *in vitro* dosisabhängig einerseits die Freisetzung von Interleukin-1, Interleukin-6 und Interferon- γ stimulierte, andererseits die durch Phytohemagglutinin stimulierte Ausschüttung von Interferon- γ und Interleukin-2 hemmte (Rosenkranz & Grundmann 1994).

1.3 Zielsetzung dieser Arbeit

Beim Einsatz eines Bakteriums zur Darreichung an Menschen sollte gewährleistet sein, dass es nicht zu Infektionen durch selbiges kommt. Ein solcher Stamm sollte möglichst nicht Träger von relevanten Pathogenitätsfaktoren sein.

Um eine genauere Aussage bezüglich der Pathogenitätsfaktoren von *E. faecalis* Symbioflor 1 treffen zu können, wurde die Entscheidung getroffen, den Stamm möglichst komplett zu sequenzieren.

Hierzu sollen zunächst die Reads aus der vorhandenen Plasmid-Bibliothek mit Reads aus einer NGS-Technologie (454-Sequenzierung) zu Contigs assembliert werden.

Die Lücken zwischen den Contigs sollen daraufhin, soweit möglich, mittels PCRs geschlossen werden. Übrige Lücken sollen mithilfe einer zu erstellenden Fosmidbank geschlossen werden.

Die Sequenz des Genoms soll dann mit bioinformatischen Mitteln analysiert und annotiert werden.

Das annotierte Genom soll im Anschluss auf das Vorhandensein von Pathogenitätsfaktoren hin untersucht werden.

Zudem soll eine komparative Genomanalyse mit anderen *E. faecalis* – Stämmen, deren Genomsequenzen vorliegen, erfolgen.

2 Material

2.1 Stämme

1. *E. faecalis* Symbioflor 1: Die vollständige Bezeichnung des Stammes lautet *Enterococcus faecalis* Symbioflor 1-02-SA, registriert unter der Nummer DSM 16431 in der Leibniz Institut DSMZ-Deutschen Sammlung Mikroorganismen und Zellkulturen GmbH (DSMZ), Inhoffen-Straße 7B, 38124 Braunschweig. Es handelt sich hierbei um einen der 10 *E. faecalis*-Klone, die im Medikament Symbioflor 1 enthalten sind. Als Kurzform wird in dieser Arbeit die Bezeichnung *E. faecalis* Symbioflor 1 verwendet.
2. EPI300-T1^R Phage T1-resistant *E.coli*: Bestandteil des CopyControl Fosmid Library Production Kits von Epicentre

2.2 Kulturmedien

- Luria-Bertani (Lysogeny Broth, LB) Bouillon: 10 g Trypton, 5 g Hefeextrakt, 10 g NaCl, 1 l aqua dest
- LB-Agar: 10 g Trypton, 5 g Hefeextrakt, 10 g NaCl, 15 g Agar, 1 l aqua dest
- Tryptic Soy Broth (TSB): Becton, Dickinson
- Genbank-Medium: 10 g Trypton, 5 g Hefe, 5 g NaCl, 6,3 g K₂HPO₄, 1,8 g KH₂PO₄, 0,45 g Na-Citrat, 0,09 g MgSO₄, 0,9 g (NH₄)₂SO, 44 g Glycerin, mit aqua dest. auf 1 l auffüllen

2.3 Chemikalien und Materialien

- 1 kb Plus DNA Ladder: Invitrogen
- Agarose: Invitrogen, Roth
- B&D I: 50 mM Glucose, 25 mM Tris Cl pH 8,0, 10 mM EDTA
- B&D II: 0,2 N NaOH, 1% SDS (Lösung täglich frisch ansetzen)
- B&D III: 60 ml 5 M Kalium-Acetat, 11,5 ml Eisessig, 28,5 ml H₂O
- Chloramphenicol: 25 mg/ml in 80%igem Ethanol, Lagerung bei -20°C
- Expand High Fidelity PCR: Roche
- Ficoll: Ficoll 400 Amersham Pharmacia Biotech AB

- Ficoll-Auftragspuffer: 1,25 g Ficoll in 7,5 ml aqua dest. über Nacht lösen, Spatelspitze Bromphenolblau zugeben. Von dieser Lösung Aliquots ansetzen, die 6:4 mit aqua dest. verdünnt werden.
- GES: 20 ml 0,5M EDTA (pH 8), 20 ml H₂O, 60 g Guanidiniumthiocyanat bei 65°C lösen, dann Lösung abkühlen. 5ml 10%iges Sarkosyl hinzufügen (Stock-Lösung 0,5g/5ml), mit H₂O auf 100ml Endvolumen auffüllen, mit 0,45 µm Filter sterilfiltrieren. Lichtgeschützt bei Raumtemperatur aufbewahren.
- Low Melting Point Agarose: Life Technologies
- Low Range Marker: New England Biolabs
- Manual PCR Extender System: 5Prime
- PCR-Puffer: Domann-Puffer (Eugen Domann 1992)
- PCR-Sequenzierung: AGOWA Gesellschaft für molekularbiologische Technologie mbH, Ostendstr. 25, 12459 Berlin
- Phusion Flash High Fidelity PCR System: Finnzymes
- Primer (zur Verwendung wurden Stocklösungen mit 20 pmol/µl angesetzt):
 - Eurofins MWG Operon, Anzingerstraße 7a, 85560 Ebersberg
 - OPERON Biotechnologies, Nattermannallee 1, 50829 Köln
- Pulsefield-certified Agarose: Biorad
- QIAGEN Puffer P1
- QIAquick PCR Purification Kit: QIAGEN
- Reaktionsgefäße: Eppendorf
- TAE-Puffer (50X): 242 g Tris, 57,1 ml Eisessig, 100 ml 0,5 M EDTA, mit aqua dest. auf 1 l auffüllen
- Taq-Polymerase: im Institut hergestellt
- TBE-Puffer (10X): 108 g Tris, 55 g Borsäure, 20 ml 0,5 M EDTA (pH 8,3), mit aqua dest. auf 1 l auffüllen
- TE-Puffer (10X): 100 mM Tris/HCl (pH 8.0), 10 mM EDTA
- TNE-Puffer: 10 mM Tris HCl (pH 7,5), 100 mM NaCl, 1mM EDTA

2.4 Geräte

- Brutschrank: Heraeus
- Elektrophoresekammer: eine von der Institutswerkstatt hergestellte Kammer

- Geldokumentationsgerät: Biorad Universal Hood II
- Hamilton-Spritze: Microliter #705
- Kamera: HTC Hermes
- Nanodrop: Nanodrop 1000 Spectrophotometer, Thermoscientific
- Pikocycler: Finnzymes
- Pipetten: Biohit, Eppendorf Research
- Pulsfeld-Gelelektrophorese: CHEF-DR II, Bio-Rad
- Speedvac: Thermolabsystems
- Thermo Cycler (PCR-Maschine): Applied Biosystems
- Tube Rotator: Kisker
- Zentrifugen: Eppendorf, Heraeus

2.5 Software

2.5.1 Lokal ausführbare Software

- Genomeviz (Ghai et al. 2004)
- GS *de novo* Assembler: 454 life sciences, Roche
- MAUVE (Darling et al. 2004)
- MEGA5 (Tamura et al. 2011)
- Mendeley Desktop
- Microsoft Office 2007
- minimap (im Institut entwickeltes Programm)
- Mugsy (Angiuoli & Salzberg 2011)
- R
- Seqman: DNASTar

2.5.2 Webbasierte Software & Ressourcen

- Augur (André Billion et al. 2006)
- CRISPRFinder (Grissa et al. 2007)
- GeCo (C. T. Kuenne et al. 2007)
- GenDB (F. Meyer 2003)
- GenomeViz (Ghai et al. 2004)

- MEGA5 (Tamura et al. 2011)
- Multi Locus Sequence Typing (<http://efaecalis.mlst.net/>)
- NCBI Blast: <http://blast.ncbi.nlm.nih.gov/> (Altschul et al. 1990; Camacho et al. 2009)
- NCBI Complete Microbial genomes
(<http://www.ncbi.nlm.nih.gov/genome/browse/>)
- Pfam (Punta et al. 2012)
- PHAST (Zhou et al. 2011)
- Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>)
- RAST (Aziz et al. 2008)
- ssgenefinder (Ho et al. 2011)

3 Methoden

3.1 Sequenzierstrategie für *E. faecalis* Symbioflor 1

Als Ausgangspunkt nach der Sanger-Sequenzierung der Plasmidbank waren zunächst 702 Contigs vorhanden. Diese Zahl war so hoch, dass ein Lückenschluss-Verfahren alleine mittels PCR einen so hohen Aufwand bedeutete, dass zunächst eine erneute Sequenzierung des gesamten Genoms mittels eines Hochdurchsatzverfahrens sinnvoll erschien (Domann et al. 2007).

Zur Komplettierung des Genoms und zur Erhöhung der Abdeckung wurde als nächstes Verfahren die Sequenzierung mittels 454-Pyrosequenzierung, einer Next-Generation-Sequencing-Technologie, angewendet.

Im Gegensatz zum Sanger-Verfahren wird hierbei die Verlängerung des DNA-Strangs durch die DNA-Polymerase während des eigentlichen Vorgangs nachgewiesen („sequencing-by-synthesis“). Beim Anhängen eines Nukleotid-Triphosphates an die wachsende Nukleotid-Kette wird ein Pyrophosphat freigesetzt. Dieses Pyrophosphat wird indirekt über eine Luciferase dazu benutzt, einen Lichtblitz zu erzeugen, welcher von einem Sensor nachgewiesen werden kann.

Die Reaktionsmischung wird so gesteuert, dass zu einem bestimmten Zeitpunkt immer nur eine Sorte an Nukleotid zur Kettenverlängerung vorhanden ist. Ein registrierter Lichtblitz kann so als die Verlängerung der Kette um eben jenes Nukleotid interpretiert werden.

Nacheinander werden so die vier Nukleotide einzeln beigemischt und wieder entfernt und die Kette verlängert.

Um eine große Menge an DNA zu sequenzieren, wird folgendermaßen vorgegangen: Die chromosomale DNA wird durch Nebulisation oder Ultraschall in Stücke von wenigen hundert Basenpaaren fragmentiert. Die Enden der Stücke werden unspezifisch an Adapter-Oligonukleotide ligiert. Die DNA wird dann denaturiert und die einzelnen Fragmente werden einzeln über den Adapter an eine komplementäre Sequenz an Kügelchen (in Nanoformat) gebunden. In einer Emulsions-PCR wird dieses DNA-Fragment in einem Wasser-in-Öl-Mikrovesikel amplifiziert. Somit wird je ein Kügelchen von vielen Amplifikaten eines DNA-Fragmentes beladen. Die

Kügelchen werden daraufhin in eine Picotiterplatte gespült, je ein Kügelchen pro Vertiefung.

In den Vertiefungen findet dann die oben genannte Sequenzierreaktion statt. In jeder Vertiefung wird somit ein anderes DNA-Fragment amplifiziert und über die Lichtreaktion der eingebauten Basen wird dessen Sequenz detektiert. Die Sequenz wird als Read ausgegeben.

Die so erhaltenen Reads (im Rahmen dieser Arbeit auch 454-Reads genannt nach der Firma, welche besagte Technologie kommerziell zur Verfügung stellt) sind im Schnitt deutlich kürzer als Sanger-Reads, zu der Zeit des Projektes etwa zwischen 200 und 250 bp. Im Gegensatz zum Sanger-Verfahren können jedoch mit verhältnismäßig geringem Aufwand sehr viele Reads generiert werden.

Das Verfahren eignet sich dadurch besonders gut zur Erhöhung der Abdeckung des Genoms, das bedeutet, dass für jede Base des Genoms im Durchschnitt mehr Reads zur Verfügung stehen und somit die Reliabilität der Consensus-Sequenz steigt.

Im Rahmen einer *de novo* - Assemblierung eines Genoms ist die relative Kürze der Reads allerdings eine Schwierigkeit für Assemblierungs-Software, da Reads umso eindeutiger mit anderen überlappt und damit assembliert werden können, je länger die Reads selber und damit die möglichen Überlappungszonen sind (Ronaghi 2001; Voelkerding et al. 2009).

Zunächst wurde die genomische DNA von *Enterococcus faecalis* Symbioflor S1 für einen Lauf mit dem 454-Sequencer verwendet. Die 454-Reads wurden dann zusammen mit den Sanger-Reads *de novo* assembliert (sog. "Hybrid-Assembly"). Hierfür wird die Software Newbler verwendet (Genome Sequencer Software, von Roche 454 life sciences zur Verfügung gestellt).

Wird ein unbekanntes Genom sequenziert, stellt sich nun die Frage, in welcher Reihenfolge die Contigs aneinandergereiht werden müssen. Bei einem vollständig neuen, unbekannten Organismus müssten dazu unter Umständen probeweise PCRs zwischen allen möglichen Anordnungen angesetzt werden.

Bei der Sequenzierung des Genoms von *Enterococcus faecalis* Symbioflor 1 bestand der Vorteil, dass ein vermutlich relativ nahe verwandter Stamm vorlag. Das Genom von *E. faecalis* V583 wurde komplett sequenziert und wird online vom NCBI zur Verfügung gestellt. (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). Da eine

vergleichsweise enge Verwandtschaft zwischen den beiden Stämmen vermutet werden konnte, konnte die Sequenz als Referenz beim Mapping der Contigs dienen. (Shankar et al. 2002)

Die assemblierten Contigs wurden mit dem Programm MAUVE an dem Referenzstamm *E. faecalis* V583 bezüglich ihrer Reihenfolge und Ausrichtung sortiert (so genanntes "Mapping"). Ausrichtung bedeutet in diesem Zusammenhang, dass für jedes Contig eine entsprechend homologe Sequenz im Referenzgenom gesucht wird, und dass die Contigs dann in der gleichen Reihenfolge aneinandergereiht werden wie ihre Entsprechungen im Referenzgenom. Hierbei müssen auch die Contigs revers komplementiert werden, deren Sequenz im Vergleich zur Referenzsequenz auf dem anderen DNA-Strang der Doppelhelix liegt (Darling et al. 2004).

Die zwischen den Contigs liegenden Lücken sollten dann mittels PCR geschlossen werden. Dazu wurden für jede Lücke 2 Primer gewählt, die am Rand der entsprechenden Contigs binden. Mit den Primern wurde ein die Lücke überspannendes PCR-Produkt generiert und ebenfalls mit diesen Primern von beiden Seiten sequenziert. Im Idealfall war damit die Lücke durch Überlappung dieser Sequenzen geschlossen. Überlappten diese PCR-Reads nicht, weil die Lücke zu groß war, mussten neue Primer am Rand der neuen Sequenzen gewählt werden, um so weiter in die Lücke vorzustoßen. Der Vorgang wurde solange wiederholt, bis die Lücke geschlossen war (so genanntes Primerwalking).

Mittels PCRs konnten so Lücken bis zu etwa 10 kbp geschlossen werden. Manche Lücken waren zu groß, um mittels PCR geschlossen werden zu können.

Für diese Fälle wurde deshalb zusätzlich eine Fosmid-Bibliothek von *E. faecalis* Symbioflor 1 erstellt. Die Bibliothek wurde ähnlich wie die Plasmid-Bibliothek durch klonieren von zufällig gescherter DNA generiert. Der Unterschied zur Plasmid-Bibliothek lag im Rahmen des Projekts vor allem in der deutlich größeren Länge der Inserts, die in den Fosmiden aufgenommen werden können. Diese können bis 40 kbp lang sein. Somit erhöht sich die Chance, in einem Fosmid die komplette unbekannte Sequenz einer Lücke vorliegen zu haben, inklusive Teile der an die Lücke anschließenden Contigs. Indem man die Fosmid-Inserts ansequenziert, versucht man Informationen über die Reihenfolge der Contigs zueinander zu

gewinnen. Vor allem bei Sequenzen, die anhand des Referenzgenoms nicht ausgerichtet werden können, kann diese Zuordnung über ein Fosmid hilfreich sein. Ebenso kann ein Fosmid mittels Primerwalking sequenziert werden, wenn es die komplette Sequenz einer Lücke enthält.

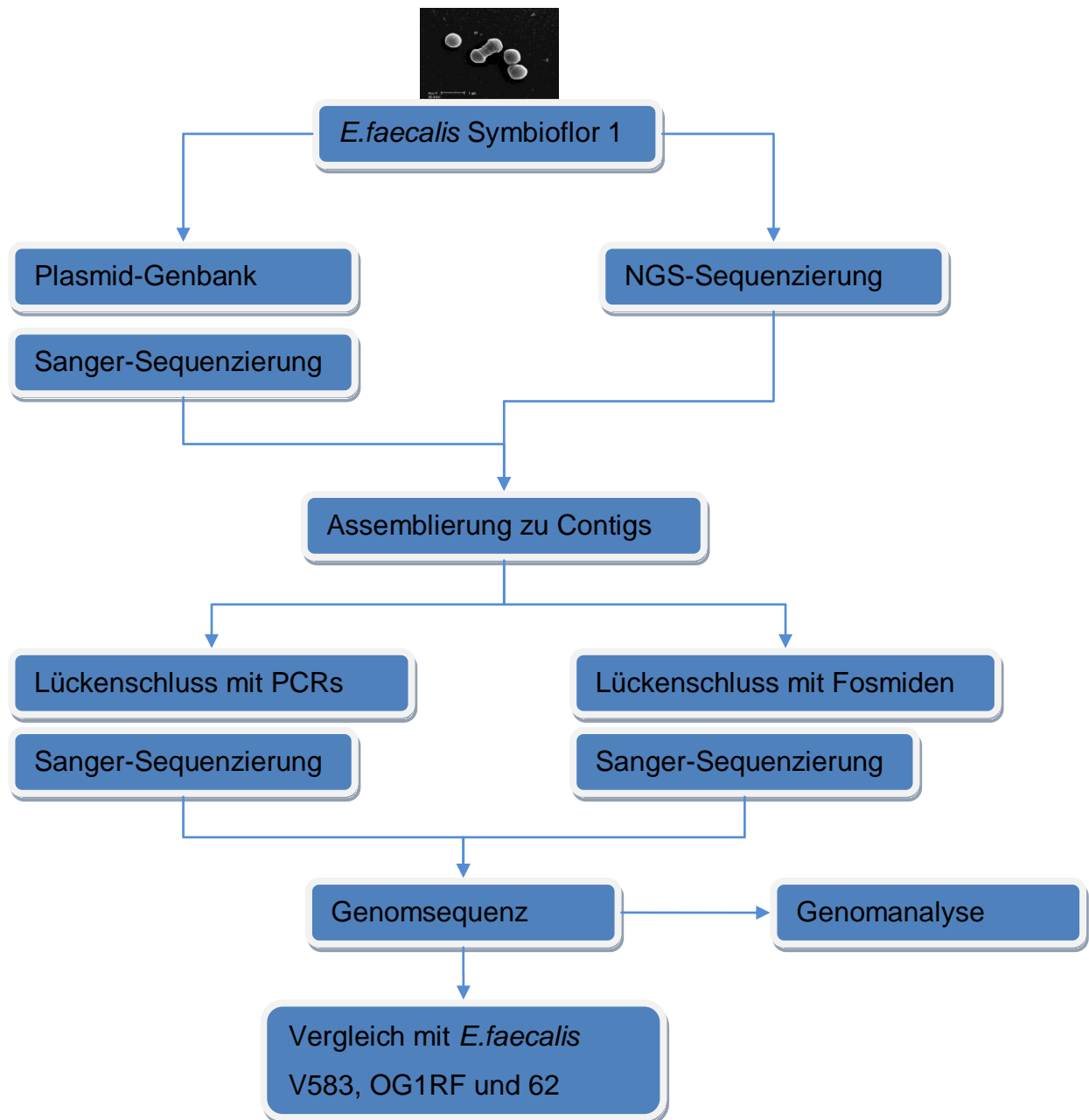


Abbildung 2: Strategie zur Zielsetzung

3.2 Isolierung chromosomaler DNA gemäß Pitcher

Diese Methode eignet sich gut zur Gewinnung chromosomaler DNA aus Übernachtskulturen von Bakterien zur Verwendung in PCRs. Die Ausbeute ist normalerweise recht hoch, allerdings ist mit Verunreinigung durch RNA zu rechnen. Verwendet wurde die Methode in erster Linie, um genug Template-DNA zu gewinnen, um die Lückenschluss-PCRs für *E. faecalis* Symbioflor S1 durchzuführen. Desweiteren wurde die DNA für den Lauf der 454-Sequenzierung ebenfalls mit dieser Methode gewonnen. (Pitcher et al. 1989)

3.2.1 Materialien

LB-Medium, 0,85%ige NaCl-Lösung, 1X TE-Puffer, RNase (DNase-frei, von Boehringer), 7,5 M NH₄-Acetat (auf 4 °C gekühlt), GES (5 mol/l Guanidiumthiocyanat (Sigma), 100 mmol/l EDTA, 0,5% (v/v) Sarkosyl (N-Lauroylsarcosinate), Lysozym (50 mg/ml in 1X TE-Puffer), Phenol-Chloroform-Isoamylalkohol im Verhältnis 25:24:1 (Sigma), Isopropanol, 70%iger Ethanol (gelagert bei -20°C), abgeschnittene Pipettenspitzen (herkömmliche 1 ml-Pipettenspitzen werden um circa 2 mm an der Spitze durch Abschneiden verkürzt um die Öffnung zu vergrößern und dann autoklaviert).

3.2.2 Protokoll

Tag 0: Eine bakterielle Übernachtskultur wird in LB-Medium angesetzt und bei 37°C mit 180 rpm über Nacht auf dem Schüttler inkubiert. Für *E. faecalis* Symbioflor S1 wurde stattdessen TSB- Medium verwendet.

Tag 1: Die Übernachtskultur wurde auf Wachstum überprüft. Die Trübung des Mediums konnte im Zweifel mithilfe einer photometrischen Messung überprüft werden. Die optische Dichte bei 600 nm sollte relativ zum Medium etwa 2,0 bis 2,2 sein.

Von der Kultur wurden 1,5 ml in ein Cap gefüllt.

Dieses wurde mit 4.000 rpm für 5 min zentrifugiert. Der Überstand wurde verworfen.

Das Pellet wurde in 0,5 ml 0,85%iger NaCl-Lösung resuspendiert. Um das Pellet besser aufzulösen empfahl es sich, das Cap mit der Unterseite auf einem Gitter, wie beispielsweise einem Reagenzglasständer, zu reiben – auch „Waschbrett-Methode“ genannt.

Anschließend wurde wieder mit 12.000 rpm für 5 min zentrifugiert. Der Überstand wurde vorsichtig mit einer Pipette abgesaugt und verworfen.

Das Pellet wurde in 100 µl Lysozym-Lösung resuspendiert (50 mg Lysozym pro ml in 1X TE), wieder mit der Waschbrett-Methode. Danach wurde erst für 30 min bei 37°C inkubiert, dann für 5 min bei Raumtemperatur. Die Hydrolyse der Zellwand mit Lysozym wird nur für grampositive Bakterien angewandt. Bei gramnegativen Bakterien würde der Schritt übersprungen werden.

Es wurde 1 µl RNase (Boehringer) hinzugegeben und dann 5 min bei Raumtemperatur inkubiert.

Es wurden 500 µl GES zugegeben. Das Cap wurde 5 mal invertiert. Dann wurde es zunächst 5 min bei Raumtemperatur inkubiert, dann 2 min auf Eis.

Es wurden 250 µl 7,5M NH₄-Acetat (4°C) zugegeben, dann wurde 4 mal invertiert und 10 min auf Eis inkubiert.

Es wurden 500 µl Phenol-Chloroform-Isoamylalkohol hinzugefügt und dann 15 min mit dem Tube Rotator langsam invertiert.

Danach wurde mit 14.000 rpm für 15 min bei 4°C zentrifugiert.

Nun wurde die obere Phase des Gemischs in ein neues Cap pipettiert. Dazu wurden die abgeschnittenen Pipettenspitzen verwendet. Mit diesen Spitzen konnte die gelöste DNA sorgfältig und vorsichtig in das frische Cap pipettiert werden, ohne dass sie geschert wurde.

Nun wurde die DNA ausgefällt – bei mehreren Ansätzen sollten die folgenden Schritte für jeden Ansatz einzeln direkt nacheinander durchgeführt werden:

Es wurde ½ Volumen Isopropanol zugefügt.

Das Cap wurde zuerst langsam in horizontaler Lage um seine Längsachse gedreht, etwa 30 sec. Dann wurde es zuerst langsam invertiert, dann immer schneller, bis es am Ende, nach weiteren 30 sec, schnell geschüttelt wurde. Durch diesen Vorgang wurde die DNA aufgewickelt und war deutlich im Cap als halb durchsichtiges Knäuel sichtbar. Anmerkung: Es gibt Bakterienstämme, deren DNA zwar bei diesem Schritt kein deutlich sichtbares Knäuel bildet, aber dennoch vorhanden und nach Ende der Aufreinigung genauso gut verwendbar ist.

Das Cap wurde nun mit 14.000 rpm für 5 min bei 4°C zentrifugiert. Der Überstand wurde verworfen.

Die DNA wurde nun mit Ethanol gewaschen. Empfohlen waren 5 Wiederholungen: Zuerst wurde 1 ml kalter 70%iger Ethanol zugegeben und das Gefäß sanft geschüttelt.

Dann wurde mit 14.000 rpm für 5 min bei 4°C zentrifugiert. Der Überstand wurde verworfen.

Beim letzten Waschgang wurde das Cap mit seinem Rand schräg an einen Zellstopf-Tupfer gepresst, um mit diesem den Rest des Alkohols abzusaugen.

Das Pellet wurde weitere 10 min in der Speedvac getrocknet.

Zuletzt wurde die DNA in 100 µl H₂O gelöst und über Nacht bei Raumtemperatur inkubiert.

Anmerkung: Die DNA tendiert dann dazu, an der Wand des Gefäßes zu kleben. Mit sanftem Schnipsen löst sie sich.

Die Ausbeute und Reinheit der isolierten DNA wurde mit dem Nanodrop bestimmt.

3.3 Polymerase-Kettenreaktion (PCR)

Dieses mittlerweile in der Molekularbiologie weit bekannte Verfahren (zurückzuführen auf Kary Mullis 1984) dient zur Vervielfältigung eines spezifischen DNA-Abschnittes: Zwei zu den Rändern des Ziel-Abschnittes komplementäre DNA-Oligonukleotide ("Primer", etwa 17-25 Nukleotide) binden an die zu amplifizierende DNA-Sequenz ("Template") und begrenzen so spezifisch den zu vervielfältigenden Abschnitt. Eine thermostabile DNA-Polymerase baut daraufhin Desoxyribonukleotide am 3'OH-Ende der Primer ein, komplementär zum jeweiligen Gegenstrang. Durch Erhitzen werden diese Stränge dann wieder voneinander getrennt. In einem neuen Zyklus binden wieder Primer (diese sind ebenso wie die Desoxyribonukleotide im Überschuss zur Reaktion zuzugeben) an ihre komplementäre Sequenz, wieder wird durch die Polymerase ein komplementärer Strang synthetisiert, wieder wird durch Erhitzen getrennt. Durch vielfaches Wiederholen dieses Zyklus – etwa 20 bis 50-fach – wird die bestimmte DNA-Sequenz exponentiell vervielfacht.

Ein üblicher Zyklus besteht demnach aus Denaturierung (Erhitzen auf 94°C), Anheften der Nukleotide ("Annealing", diese Temperatur hängt von der Sequenz und Länge der Primer ab und liegt üblicherweise zwischen 50 und 60°C) und Extension (Arbeitstemperatur der Polymerase, bei üblicher *Taq*-Polymerase 72°C).

Als Reagenzien werden benötigt: Template-DNA, die kopiert werden soll, Primer, die die Zielsequenz begrenzen, Desoxyribonukleotid-Triphosphate als Bausteine zur Strangverlängerung, eine hitzestabile DNA-Polymerase und ein zur Polymerase passender Puffer. Weiterhin wird eine PCR-Maschine benötigt, die die Reaktion nach Programm auf die benötigten Temperaturen erhitzt und abkühlt (Löffler et al. 2002).

Die Programme wurden bezüglich Temperaturen und Zeiten von den Hersteller-Protokollen der verwendeten PCR-Systeme übernommen, allerdings musste die Annealing-Temperatur an die jeweils verwendeten Primer angepasst werden. Für die Primer wurde eine Schmelztemperatur berechnet bzw. vom Hersteller genannt. Die Annealing-Temperatur wurde zunächst etwa 3 bis 4°C unter der niedrigeren Schmelztemperatur der beiden Primer angesetzt. Eine zu hohe Annealing-Temperatur verhindert die Ausbildung von PCR-Produkten, eine zu niedrige Temperatur führt unter Umständen dazu, dass die Primer an Stellen binden, die nicht 100%ig komplementär sind, dadurch können falsche Zielsequenzen amplifiziert werden: das PCR-Produkt wird unspezifisch.

Im Rahmen der Lückenschluss-PCRs dieses Projektes wurde isolierte chromosomale DNA von *E. faecalis* Symbioflor 1 als Template verwendet.

Die Lücken im Assembly enthalten oft repetitive Elemente, also Sequenzen, die gleich oder sehr ähnlich an verschiedenen Stellen im Genom vorkommen. Da die gesamte DNA des Stammes als Template dient, haben PCRs dieser Lücken oft unspezifische Amplifikate geliefert. Wenn eine Erhöhung der Annealing-Temperatur dies nicht behob, mussten neue Primer kreiert werden.

3.3.1 Expand High Fidelity PCR (Roche)

Dieses kommerzielle PCR-Kit von Roche wurde hauptsächlich dazu benutzt, die PCRs für den Lückenschluss anzusetzen. Das Kit besteht aus einem fertigen Polymerase-Mix, einem Puffer-Mix, sowie optional einem Puffer-Mix ohne Mg^{2+} und einer Mg^{2+} -Lösung. Die Polymerasen sind eine *Tgo* DNA Polymerase mit inherenter 3'-5'-Exonuklease-Aktivität und eine thermostabile *Taq*-Polymerase.

Das PCR-System ist laut Herstellerangaben dazu optimiert, Fragmente bis zu 5.000 bp effizient und mit einer 3-fach niedrigeren Fehlerrate als mit *Taq*-Polymerase allein

zu amplifizieren. Es sind auch PCRs bis zu 9.000 bp und mehr möglich (Barnes 1994).

Das System zeigte sich zuverlässig und war in der Lage, die meisten für den Lückenschluss benötigten Amplifikate herzustellen.

Die PCR-Ansätze wurden nach Herstellerangaben angesetzt. Als Template für die Lückenschluss-PCRs diente die chromosomale DNA von *E. faecalis* Symbioflor 1, die mit der modifizierten Methode nach Pitcher isoliert wurde.

Folgendes Standard-Protokoll wurde für die Reaktion benutzt.

94°C	4 min		
94°C		15 sec	10 Zyklen
50°C*		30 sec	
68°C	8 min		
94°C		15 sec	20 Zyklen
50°C*		30 sec	
68°C	8 min	+5 sec pro Zyklus	
72°C	7 min		
4°C	dauerhaft		

*) Annealing Temperatur, wird an die verwendeten Primer angepasst

3.3.2 Manual PCR Extender System (5Prime)

Dieses ist ein PCR-Kit der Firma 5Prime, welches früher unter dem Namen "TripleMaster" vertrieben wurde. Es kam zum Einsatz, wenn das Kit von Roche keine Ergebnisse bei den Lückenschluss-PCRs lieferte. Für manche Lücken war nur jeweils eins dieser PCR-Kits in der Lage, PCR-Produkte zu produzieren.

Auch dieses Kit beinhaltet eine Mischung verschiedener Polymerasen, die laut Herstellerangaben für Amplifikate von bis zu 50 kbp geeignet seien und über eine sehr niedrige Fehlerrate verfügen, zudem seien sie auch in der Lage schwierige, GC-haltige Templates zu amplifizieren. Es werden zwei verschiedene Puffer mitgeliefert, die sich für unterschiedliche Arten Template und unterschiedliche Längen der Amplifikate eignen.

Das Kit wurde mit beiden Puffern nach Angaben des Herstellers verwendet. Das PCR-Programm wurde gegebenenfalls an die Annealing-Temperatur für die verwendeten Primer angepasst.

3.3.3 Phusion Flash High Fidelity PCR System (Finnzymes)

Ein weiteres PCR-System, das bei schwierigen Templates Verwendung fand. Auch dieses ist laut Herstellerangaben gut geeignet für schwierige Templates. Eine Besonderheit des Systems ist, dass es mit einer dafür vorgesehenen PCR-Maschine, dem Pikocycler, sehr schnelle Protokolle fahren kann – bis 15 sec / 1 kb). Das System wurde nach Herstellerangaben angewendet. Die Annealing-Temperatur wurde auch hier wieder an das jeweils verwendete Primerpaar angepasst.

3.3.4 Aufreinigung von PCR-Produkten mittels QIAquick PCR Purification Kit

Die PCR-Produkte, die sequenziert werden sollten, wurden mit Hilfe dieses Kits aufgereinigt. Das System funktioniert mit Säulchen, an deren Silikon-Membran die DNA gebunden wird, wenn hoher Salzgehalt herrscht. Die so gebundene DNA wird zunächst gereinigt, indem Waschpuffer durch Zentrifugation an ihr vorbeigespült wird. Danach wird die DNA mit (salzfreiem) Wasser von der Membran gelöst und durch Zentrifugation aus dem Säulchen gespült. Die Aufreinigung wurde nach dem Protokoll des Herstellers durchgeführt.

3.3.5 Sequenzierung der PCR-Produkte

Die aufgereinigten PCR-Produkte wurden von der Firma AGOWA sequenziert.

Hierzu wurden jeweils 12 µl PCR-Produkt mit H₂O auf die Konzentration eingestellt, die von der Firma für die jeweilige Größe des PCR-Produktes empfohlen wurde. Für größere Produkte wurde eine höhere DNA-Konzentration empfohlen (zwischen 10-40 ng/µl).

Um die Konzentration zu bestimmen, wurden 1,6 µl PCR-Produkt auf dem Nanodrop gemessen und das PCR-Produkt dann entsprechend den Vorgaben mit H₂O verdünnt.

Dazu wurden 2 µl von dem Primer zugegeben, mit dem sequenziert werden soll. Im Unterschied zur PCR wird bei der Sequenzier-Reaktion nur 1 Primer verwendet.

In der Regel wurden so für jedes PCR-Produkt zwei Ansätze angesetzt, je einen für jeden der beiden PCR-Primer.

Wenn ein PCR-Produkt größer als ungefähr 2.000 bp war, reichten die Reads (etwa 1.000 bp), die mit den PCR-Primern gewonnen wurden, meistens nicht ganz aus, das komplette PCR-Produkt komplett zu sequenzieren. In diesen Fällen wurden basierend auf diesen Reads weitere Primer entworfen, die weiter innen im PCR-Produkt banden, und das PCR-Produkt wurde dann erneut mit jenen Primern sequenziert.

Die resultierenden Reads wurden von AGOWA online als *.ab1-Dateien zur Verfügung gestellt.

3.4 Gelelektrophorese von DNA-Fragmenten

Die Elektrophorese von DNA dient dazu, Fragmente nach ihrer Länge aufzutrennen. Es wird ein Agarosegel gegossen, an dessen oberem Ende die DNA in Taschen gefüllt wird. Dieses Gel wird in eine Kammer mit einem Puffer gelegt, und von außen wird über einen definierten Zeitraum elektrische Spannung angelegt. Bedingt durch ihre negative Ladung durch Phosphat werden die DNA-Fragmente durch das Gel zur Anode gezogen. Das Agarosegel bremst hierbei längere Fragmente stärker als kürzere, daher wandern letztere schneller. Parallel zu den zu analysierenden Fragmenten wird üblicherweise ein Marker aufgetragen, der aus einer Mischung aus DNA-Fragmenten definierter Längen besteht. Das Gel wird entweder vor dem Lauf oder hinterher mit einem Farbstoff für DNA angefärbt, dann wird mit einem Geldokumentationsgerät ein Foto gemacht. Für dieses Projekt wurde das Geldokumentationsgerät von Biorad verwendet.

Zudem kann die Gelelektrophorese auch schlicht verwendet werden, DNA semiquantitativ darzustellen. Vergleichbare Mengen an DNA bilden vergleichbare Leucht-Intensitäten aus.

3.4.1 Herkömmliche Agarose-Gelelektrophorese

Dieses Verfahren wurde benutzt, um DNA-Fragmente, die zwischen 100 bp und etwas über 12 kbp lang waren, darzustellen, nachzuweisen und deren Länge zu bestimmen.

Als Hauptanwendung seien hier die PCR-Produkte der zahlreichen Lückenschluss-PCRs genannt.

Zunächst wurde Agarose in 1X-TBE-Puffer gelöst und in der Mikrowelle aufgeköcht. Je nach zu erwartender DNA-Fragment-Länge wurde eine Agarose-Konzentration zwischen 0,7 und 1% gewählt – für die Standard-PCR-Fragmente 1%. Für größere

Fragmente wurde die Konzentration verringert, auch wurde statt TBE-Puffer hier teilweise TAE-Puffer verwendet.

Nachdem die Lösung wieder etwas abgekühlt, aber immer noch flüssig war, wurde Ethidiumbromid hinzugegeben (0,5 ng/ml) und vorsichtig geschwenkt – es musste unter dem Abzug gearbeitet werden. Dann wurde die Lösung in den gut abgedichteten Gel-Schlitten gegossen. Nachdem das Gel abgekühlt und fest geworden war, wurde es in die Laufkammer gelegt und die Kämme, die die Taschen für die Proben bilden, wurden entfernt. Die Laufkammer wurde mit so viel 1X-TBE-Puffer gefüllt, dass das Gel unter der Flüssigkeitsoberfläche lag.

Je nach zu erwartender DNA-Menge wurden zwischen 2 und 10 µl Probe aufgetragen. Hierzu wurde die aufzutragende Menge vorher mit dem gleichen Volumen Ficoll-Auftragspuffer gemischt und dann in die Taschen des Gels pipettiert.

Als Marker diente die 1Kb Plus DNA Ladder von Invitrogen.

Je nach zu erwartender Länge der Fragmente wurde die Elektrophorese für 25 bis 60 min mit 120 V bis 150 V gestartet. Die Stromstärke wurde auf Maximum gesetzt. Für längere Fragmente wurde eine längere Laufzeit gewählt. Die Erhöhung der Spannung erhöhte die Geschwindigkeit der DNA-Fragmente.

3.4.2 Pulsfeld-Gelelektrophorese (PFGE)

Die herkömmliche Agarose-Gelelektrophorese trennt Fragmente ab einer Größe von mehr als 12 kbp nur noch unzureichend scharf voneinander. Für größere Fragmente, wie sie zur Erstellung der Fosmidbank benötigt wurden, wurde daher die Pulsfeld-Gelelektrophorese verwendet. Bei dieser Variante wechselt kontinuierlich die Richtung des elektrischen Feldes, sodass längere DNA-Fragmente leichter durch das molekulare Geflecht des Agarosegels gezogen werden.

Das Gel für die PFGE wurde ähnlich wie für die herkömmliche Agarose-Gelelektrophorese vorbereitet: Es wurde eine 1%ige Agarose-Lösung in 1X TBE-Puffer aufgekocht. Hierzu wurde Pulsfeld-zertifizierte Agarose verwendet. Die heiße Agarose-Lösung wurde zunächst ein wenig abgekühlt und dann im noch flüssigen Zustand in den Gelschlitten gegossen. Danach wurde dieser zum Kühlen etwa 30 min im Kühlschrank bei 4°C belassen. Eine Färbung mit Ethidiumbromid erfolgte erst nach der Elektrophorese, um die Kammer nicht zu verunreinigen. Das feste Gel wurde mit den DNA-Proben beladen, ein Auftragspuffer wurde nicht verwendet. Die

Proben wurden mit heißer flüssiger Low-Melting-Point-Agarose in den Taschen versiegelt. Hierbei war darauf zu achten, keine Luftblasen in den Taschen einzuschließen. Danach wurde das Gel mitsamt dem Schlitten in die Elektrophoresekammer gelegt. Als Laufpuffer diente 1X TBE. Die Elektrophorese wurde mit folgenden Parametern gestartet: Laufzeit 17 h, Initial Switch 1 sec, Final Switch 12 sec, Spannung 6 V/cm.

Hinterher wurde das Gel für circa 10 min in einem Ethidiumbromid-Bad gefärbt, dann circa 10 min in 1X-TBE-Puffer entfärbt, und mit dem Biorad-Geldokumentationsgerät fotografiert und asserviert.

3.5 Erstellung einer Fosmidbibliothek von *E. faecalis* Symbioflor 1

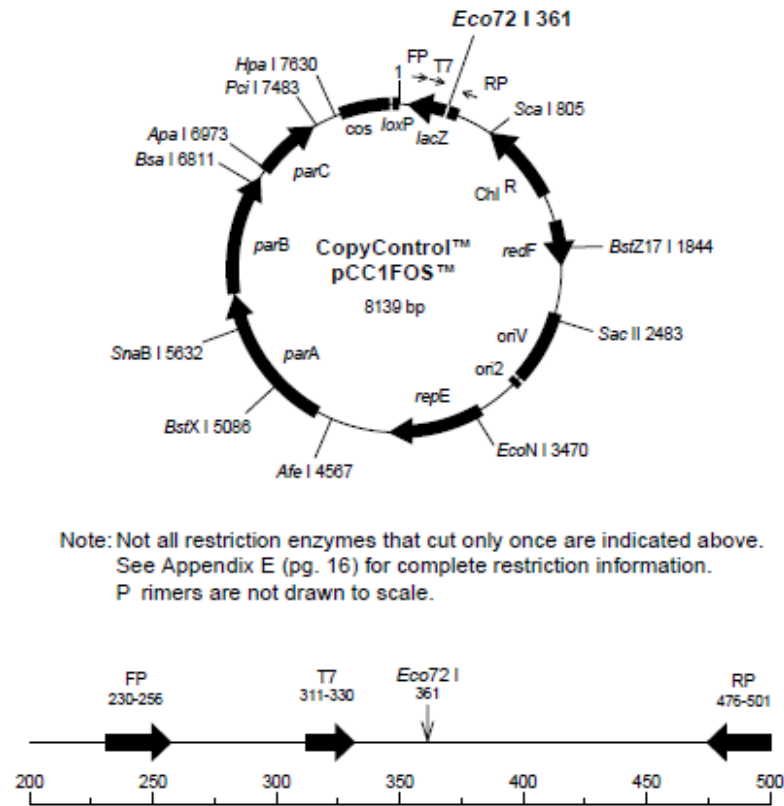
3.5.1 Herstellung der Fosmid-Bibliothek

Die Fosmidbibliothek wurde erstellt mit Hilfe des CopyControl Fosmid Library Production Kit von Epicentre.

Es dient dazu, DNA-Fragmente von etwa 40.000 bp als Insert in einen Fosmid-Vektor zu ligieren, welcher dann mittels Phagen in einen Bakterienstamm eingeschleust wird (Transduktion). Bakterienklone, die den Vektor samt Insert enthalten, werden selektiert, in Medium angezüchtet und für weitere Anwendungen in Mikrotiter-Platten eingefroren.

Für *E. faecalis* Symbioflor 1 wurde dieses Verfahren gewählt, um größere Fragmente des Genoms gezielt sequenzieren zu können. Insbesondere sollten die Fosmid-Inserts dazu dienen, Aussagen über Lücken im Genom zu erhalten, wo keine PCR vorlag.

Der benutzte Vektor heit pCC1Fos:



FP = pCC1TM/pEpiFOSTM Foward Sequencing Primer 5' GGATGTGCTGCAAGGCGATTAAGTTGG 3'
 RP = pCC1TM/pEpiFOSTM Reverse Sequencing Primer 5' CTCGTATGTTGTGTGGAATTGTGAGC 3'
 T7 = T7 Promoter Primer 5' TAATACGACTCACTATAGGG 3'

Abbildung 3: Vektor pcc1Fos

Abbildung aus dem Protokoll des Herstellers (CopyControl Fosmid Library Production Kit von Epicentre)

Der Vektor hat unter anderem folgende Eigenschaften:

- Chloramphenicol-Resistenz als Marker zur Selektion
- Sowohl einen Replikationsursprung zur Einzelkopie als auch einen zur Mehrfachkopie
- Eine cos site zur Verpackung in lambda-Phagen.

Nachdem der Vektor mit einem Insert ligiert wurde, wird er in lambda-Phagen-Partikel verpackt. Damit werden daraufhin die ebenfalls im Kit enthaltenen EPI300-T1R-*E.coli*-Bakterien infiziert. Klone, die einen Vektor mit Insert enthalten, werden nun über Chloramphenicol-haltige Agarplatten selektioniert.

Hierbei sollen nur die Klone überleben, welche den pCC1Fos-Vektor aufgenommen haben und von diesen auch nur jene, bei denen im Vektor ein Insert ligiert wurde.

Die so selektionierten Klone werden in Mikrotiter-Platten in Medium mit Glycerin zunächst nochmal über Nacht vermehrt und dann eingefroren und aufgehoben.

Durch Zugabe der von Epicentre bereitgestellten Multicopy-Induktionslösung können die Bakterien dazu gebracht werden, die Fosmid-Anzahl pro Zelle von 1 auf 10-200 zu erhöhen, um bessere Ausbeuten bei der späteren Isolierung zu erreichen. Hierzu wird die Replikation des Fosmids vom Replikationsursprung oriV gestartet. Dazu ist jedoch das Genprodukt *trfA* nötig, welches die EPI300-Bakterien nur nach Zugabe der besagten Lösung exprimieren.

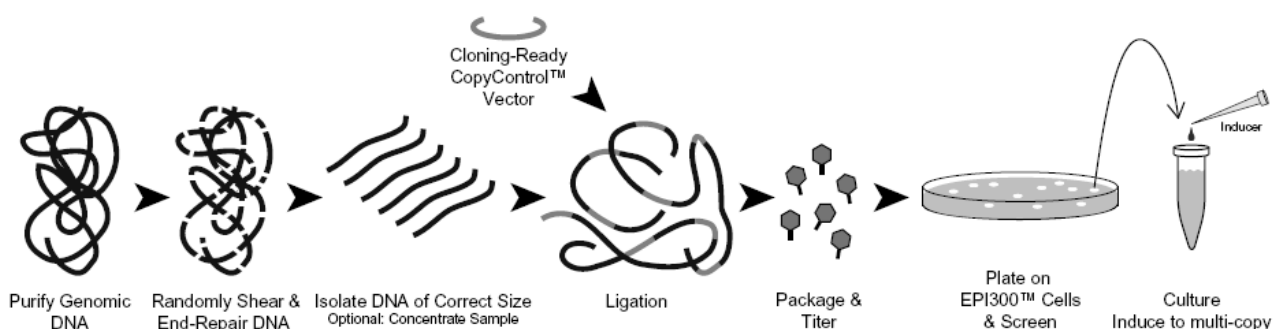


Abbildung 4: Herstellung der Fosmid-Bibliothek

Abbildung aus dem Protokoll des Herstellers (CopyControl Fosmid Library Production Kit von Epicentre)

3.5.1.1 Isolierung der DNA

Die DNA wurde mit der modifizierten Methode nach Pitcher isoliert.

3.5.1.2 Scheren der DNA

Es hat sich gezeigt, dass zum späteren Zeitpunkt eine große Menge an DNA wichtig ist. Daher wurden für die Größenselektion und Vorbereitung der DNA drei parallele Ansätze mit jeweils 50 μ l durchgeführt, jeder mit einer Anfangskonzentration von etwas über 1.200 ng/ μ l.

Jeder dieser Ansätze wurde mit der Hamilton-Spritze geschert. Dazu wurde mit einer Frequenz von etwa 100 Hz 50-mal auf und ab pipettiert.

Anmerkung: Das Scheren der DNA sollte für die entsprechende Größe optimiert werden. Hierzu sollte man ein paar Ansätze versuchsweise scheren und mit der unten genannten Pulsfeldgelelektrophorese überprüfen. Idealerweise sollte die meiste DNA auf der Höhe von 40 kbp liegen. So wie dieser Versuchs-Ansatz geschert wurde, sollte dann auch der für die Klonierung vorgesehene Ansatz geschert werden.

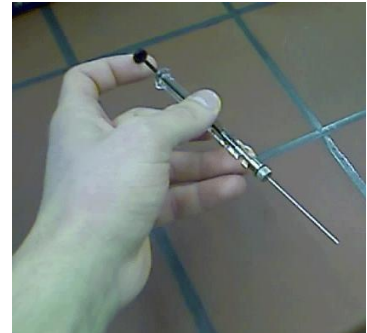


Abbildung 5: Hamilton-Spritze

Die DNA wird durch die kleine Nadelöffnung hoch und runter pipettiert und dadurch geschert

3.5.1.3 End-Repair der DNA

Hierbei werden die Enden der DNA geglättet, das heißt beide Stränge enden an der gleichen Stelle, es gibt keinen Überhang, keine so genannten "Sticky Ends" mehr. Der Vorgang wurde nach Angaben des Herstellers durchgeführt. Es wurden statt der empfohlenen 20 μ g DNA jedoch 30 μ g pro Ansatz verwendet.

3.5.1.4 Größenselektion der DNA

Um Fragmente mit der Länge von 40 kbp als Insert zu bekommen, wurde die gescherte DNA mittels Pulsfeldgelelektrophorese aufgetrennt und eine Bande bei der entsprechenden Größe ausgeschnitten.

Hierzu wurde 1% Low-Melting-Point-Agarose in 1X-TBE-Puffer gelöst und in der Mikrowelle aufgeköcht. Für jeden der drei Ansätze wurden doppelte Taschen gegossen, indem die Lücke zwischen zwei Zähnen des Gelkamms mit Klebestreifen zugeklebt wurden. Das Gel wurde nun gekühlt bis es fest war. Dann wurden zuerst die DNA-Proben in die Taschen pipettiert, danach wurden die Taschen mit heißer Agarose versiegelt. Dazu wurde die gleiche Agarose-Lösung wie für das Gel benutzt. Als Größenkontrolle diente der Low-Range-Marker. Er lag als Gel-Rolle vor – ein Stück wurde mit einem Skalpell abgeschnitten und in die Tasche geschoben, danach ebenfalls mit flüssiger Agarose versiegelt. Das Gel wurde auf der Platte, auf der es

gegossen wurde, vorsichtig in die Gelkammer gelegt. Als Laufpuffer wurde 1X-TBE verwendet. Die Elektrophorese wurde mit folgenden Parametern gestartet: Laufzeit 17 h, Initial Switch 1 sec, Final Switch 12 sec, Spannung 6 V/cm.

Das Gel wurde hinterher in einem Ethidiumbromid-Bad gefärbt. Allerdings durfte die DNA der Probe selbst nicht gefärbt werden, um den weiteren Prozess nicht zu stören. Deshalb wurden nur die Streifen, die den Marker enthielten, aus dem Gel ausgeschnitten und einzeln gefärbt. Die Streifen mit der DNA der Probe wurden solange in Saranfolie gewickelt.

Nach der Färbung wurden die Streifen mit dem Marker auf dem UV-Tisch betrachtet: Mit einem Skalpell wurde jeweils eine Markierung bei der 40 kbp-Bande des Markers ins Gel geschnitten.



Abbildung 6: Größenselektion der DNA

Links: die Gelstücke auf dem UV-Tisch

Mitte: alle Gelstücke nebeneinander in Saranfolie

Rechts: Kontrolle des zusammengesetzten Gels nach Entnahme der 40 kbp Bande für das Insert

Das Gel wurde danach wieder zusammengesetzt. Aus den Gelstreifen mit der Versuchs-DNA wurde nun ein horizontaler Streifen herausgeschnitten, der auf der Höhe der Markierungen lag. In diesen Streifen sollte nun DNA mit der Länge von etwa 40 kbp enthalten sein. Die Streifen wurden in vorher ausgewogene 1,5 ml-Caps gesteckt. Dann wurde das Cap wieder gewogen – die Differenz zum Leergewicht ergab das Gewicht des Gelstreifens. Zur Kontrolle wurde der Rest des Gels zusammengesetzt und auf dem Geldokumentationsgerät fotografiert.

3.5.1.5 Gelverdau und DNA-Rückgewinnung

Der Verdau des Gels wurde nach dem Protokoll des Herstellers durchgeführt.

Die drei Ansätze wurden jeweils in 40 µl TE-Puffer resuspendiert, dann vereint. Die Konzentration wurde mit dem Nanodrop-Gerät gemessen, sie lag bei 19,7 ng/µl.

Um die Konzentration zu erhöhen und gleichzeitig den Alkohol, der sich noch in der Probe befand, abdampfen zu lassen, wurde die Probe nun 50 min in der Speedvac getrocknet. Danach wurde die Konzentration erneut mit dem Nanodrop-Gerät gemessen und betrug nun 96,0 ng/µl.

3.5.1.6 Ligation

Die Ligation wurde nach den Angaben des Herstellers durchgeführt. Dabei wurden 3 µl der oben genannten DNA-Lösung verwendet – das entsprach ~288 ng, also etwas mehr als die im Protokoll geforderten 250 ng.

3.5.1.7 Packaging

Das Packaging der Vektoren in die lambda-Phagen-Partikel wurde nach den Angaben des Herstellers durchgeführt. Die Inkubationszeit der EPI300-Bakterien aus der Übernachtskultur bis zur gewünschten OD betrug 1 h 30 min.

3.5.1.8 Infizierung der EPI300-Bakterien, Titrierung der verpackten Klone

Von den verpackten Fosmid-Klonen wurden zunächst Verdünnungen wie im Protokoll angefertigt, jedoch beginnend mit 1:10 und bis zu 1:10.000.

Für jede Verdünnung, und für die unverdünnte Lösung, wurde eine Infektion mit 10 µl auf 100 µl der vorbereiteten EPI300-Bakterien-Lösung durchgeführt. Wie im Protokoll vorgesehen wurden diese Infektionsansätze 20 min bei 37°C inkubiert, dann auf je eine LB-Platte mit 12,5 µg/ml Chloramphenicol ausplattiert und über Nacht bei 37°C inkubiert.

Weiterhin wurden auch Infektions-Ansätze mit bis zu 100 µl der unverdünnten verpackten Fosmid-Klone angesetzt.

Die unverdünnte Lösung ergab die günstigste Klondichte von etwa 100 Klonen pro Platte und wurde weiter verwendet.



Abbildung 7: Fosmid-Klone nach erfolgreicher Infizierung

3.5.1.9 Anlegen der Mikrotiter-Platten

Die Fosmid-Klone sollten nun gepickt werden, das heißt, jeder einzelne Klon sollte in einem Well einer Mikrotiterplatte konserviert werden.

Zunächst wurde Genbank-Medium mit Chloramphenicol zu einer Endkonzentration von 12,5 µg/ml versetzt.

Von diesem Medium wurden in eine 96-Well-Genbank-Mikrotiter-Platte je 200 µl pro Well pipettiert.

Nun wurde mit autoklavierten Zahnstochern von jeder einzelnen Kolonie Material gepickt und in ein Well gestreift. Jeder Zahnstocher wurde nur einmal verwendet; auch wurden Kolonien gemieden, die mit anderen Kolonien verschmolzen waren, denn jedes Well sollte nur einen einzigen Klon enthalten.

Die volle Platte wurde mit dem Deckel zugedeckt und auf ein Tablett gestellt, auf dem mit 70%igem Ethanol getränkte Zellstofftücher lagen. Das Tablett mit den Genbank-Platten wurde dann in eine Plastiktüte verpackt. Es wurde über Nacht bei 37°C inkubiert.

Die Genbank-Platten wurden beschriftet mit "EfS1 Fos *n*", wobei *n* eine fortlaufende Nummer war. Die Wells der Platte trugen Beschriftungen von 1A bis 12H; wichtig war es, bei allen weiteren Arbeiten die Orientierung der Platte beizubehalten. Über den Namen der Platte und die Position innerhalb der Platte kann jeder einzelne Klon wieder identifiziert werden.

Am folgenden Tag wurden die Platten mit Parafilm verschlossen und in einem Gefrierschrank bei -80°C verwahrt.

Bei jedem Auftauen und Einfrieren der Fosmid-Genbank-Platten gehen Bakterien verloren. Es war daher sinnvoll, eine Sicherungskopie anzulegen.

Dazu wurden wie oben beschrieben Genbank-Platten mit Genbank-Medium mit Chloramphenicol gefüllt. Anstatt die Klone einzeln von den Agarplatten zu picken, wurde die Original-Genbank-Platte verwendet: Mit einem autoklavierten, hierfür vorgesehen 48er-Metall-Stempel wurde Material aus der Original-Platte in die neue Platte übertragen; pro Genbank-Platte waren zwei solcher Stempel notwendig.

Während die Original-Platte sofort darauf wieder eingefroren wurde, wurde die Kopie analog über Nacht inkubiert, dann mit Parafilm verschlossen, und dann ebenfalls eingefroren.

Sie erhielten den gleichen Namen wie ihre entsprechenden Originale mit einem zusätzlichen „K“ für Kopie.

3.5.2 Fosmid-Restriktionsverdau

Ein Restriktionsenzym ist eine Endonuklease, welche DNA an einer spezifischen Stelle schneidet. Verschiedene Restriktionsenzyme haben unterschiedliche Ziel-Sequenzen, meist sind es Palindrome von wenigen Basenpaaren. DNA-Stücke mit bekannten Sequenzen lassen sich so an definierten Stellen schneiden, sofern man an der betreffenden Stelle eine Erkennungssequenz für ein Restriktionsenzym findet. Gegebenenfalls muss man noch sichergehen, dass die gleiche Erkennungssequenz nicht ein weiteres Mal an anderer Stelle in der DNA vorkommt, die geschnitten werden soll.

Die verschiedenen Enzyme benötigen verschiedene Puffer und zum Teil verschiedene Reaktionsbedingungen, was Temperatur und Reaktionsdauer angeht.

Um einen Überblick über die Inserts der Fosmide zu erhalten, wurde eine Stichprobe der Fosmid-Klone isoliert und mit *BamH1* verdaut, einem Restriktionsenzym, welches im Vektor pCC1Fos rechts und links neben dem Insert schneidet. Dadurch wurde das Insert aus dem Vektor gelöst.

Außerdem ist es so, dass *BamH1* auch in den Inserts schneiden kann. Die Sequenz der Inserts ist bei einer Stichprobe unbekannt, die Schnittstellen für *BamH1* sind also zufällig verteilt.

Trägt man die mit *BamH1* geschnittenen Fosmid-Isolate auf einem Agarosegel auf, sollte jedes Fosmid, solange es unterschiedliche Klone sind, durch unterschiedlich lange Fragmente seines jeweiligen Inserts ein anderes Muster bieten.

3.5.3 Screening der Fosmidbibliothek

Zur Finishing-Phase des Genomprojektes erschien es wenig ratsam, die vorhandene Fosmidbibliothek im Gesamten zu sequenzieren. Eine Sequenzierung mit vektorständigen Primern liefert Reads, die nur einen Bruchteil der im Fosmid im Schnitt enthaltenen Sequenz darstellen – Reads sind ungefähr 1.000 bp, Fosmid-Inserts etwa 40 kbp lang. In der Finishing-Phase des Genomprojektes fehlten die Sequenzen weniger Gaps. Durch das Sequenzieren einzelner oder auch vieler

zufällig bestimmter Klon Reads zu erhalten, die die bekannte Sequenz in diese Gaps hinein verlängern, erscheint unwahrscheinlich.

Daher war eine Methode erforderlich, um die Klone, die mit ihrem Insert ein vorhandenes Gap abdecken, zu identifizieren, um diese dann gezielt sequenzieren zu können. Hat man einen Fosmid-Klon identifiziert, der vermeintlich Gap-Sequenzen enthält, lassen sich durch Sequenzierung desselben mit den vektorständigen Primern unter Umständen auch Informationen darüber gewinnen, wie und in welcher Ausrichtung die Contigs um dieses Gap herum angeordnet sind.

Die verwendete Methode funktioniert mittels gepoolter PCR.

Zunächst wurden für die entsprechenden Gaps Primerpaare gesucht. Die Primer sollten ein etwa 400 bp großes Stück begrenzen, welches nahe (etwa 500 bp) an dem entsprechenden Gap, aber noch innerhalb schon bekannter Sequenzen in einem Contig lag. So wurde für jeden Rand eines Gaps ein Primerpaar ausgewählt. Die Primer wurden benannt nach dem Schema: "EfFosSc_p2_*Contigname*_(*Beg/End*)_(*F/R*)", wobei der *Contigname* der Name des aktuellen Contigs war, *Beg* oder *End* den vorderen oder hinteren Rand des Contigs bezeichnet und *F* oder *R* für forward oder reverse gewählt wurde, für die letzten beiden Affixe galt maßgebend die Ausrichtung am Referenzstamm.

Nun wurde die Fosmidbibliothek mittels dieser Primerpaare gescreent. Jeweils 6 Klone wurden in einer gemeinsamen PCR mit einem Primerpaar auf Vorliegen der entsprechenden Sequenz getestet. Bei einem positiven Ergebnis bedeutete es, dass mindestens einer der Klone die gesuchte Sequenz beinhaltet. Daraufhin wurde die PCR mit je 1 Klon pro Ansatz wiederholt. Die nun positiven Klone sollten die gesuchte Sequenz enthalten.

Die mit dieser Methode gefundenen positiven Klone wurden dann zunächst mit dem Primerpaar, was vermeintlich am anderen Ende des Gaps liegt, erneut getestet. War dies auch positiv, lag vermutlich das gesamte Gap innerhalb des Inserts dieses Fosmid-Klons.

Außerdem wurde der gefundene Fosmid-Klon aufgetaut und auf einer LB-Platte kultiviert, in einer Glycerin-Kultur erneut konserviert, und das Fosmid wurde isoliert und mit den Vektor-Primern ansequenziert. Die resultierenden Reads sollten eine Einordnung des Inserts innerhalb des Assembly ermöglichen.

Für die PCR wurde ein Mix mit hauseigen gewonnener *Taq*-Polymerase verwendet. Der Mastermix wurde folgendermaßen angesetzt und reichte für 50 Reaktionen. Er wurde nach Anfertigung auf 50 PCR-Caps verteilt und bei -20°C eingefroren.

Fosmid-Screening-Mastermix für 50 Reaktionen:

2.040 µl H₂O, 500 µl Domann-Puffer, 200 µl dNTPs (1,25 mM), 10 µl *Taq*-Polymerase. Aliquotieren zu je 95 µl.

Die vorbereiteten PCR-Ansätze wurden aufgetaut und mit je 2 µl von den beiden entsprechenden Primerlösungen versetzt, dazu kam 1 µl Wasser – ein Ansetzen eines Mastermixes mit beiden Primern und Wasser erleichterte die Arbeit.

Um Material für die PCRs zu gewinnen, wurden die tiefgefrorenen Fosmidbank-Platten zunächst auf Agarplatten gestempelt und über Nacht bei 37°C kultiviert.

Von den gewachsenen Kolonien wurde nun mittels Zahnstocher Material gepickt und in ein beschriftetes PCR-Cap mit PCR-Ansatz gestreift. Pro Agarplatte mit 48 Kolonien wurden 8 PCR-Ansätze mit je 6 Klonen angesetzt. Man sollte nicht zuviel Material nehmen, da sonst die PCR gestört wird.

Als Positiv-Kontrolle wurde als Template ein Pick von einer Kolonie des *E. faecalis* Symbioflor S1-Stammes genommen. In die Negativ-Kontrolle kam nichts zusätzlich.

Die PCR wurde nach folgendem Protokoll gestartet.

94°C	10 min		
94°C		40 sec	30 Zyklen
56°C*		40 sec	
72°C	2 min		
72°C	7 min		
4°C	dauerhaft		

*) Annealing Temperatur, wird an die verwendeten Primer angepasst

Die hohe Temperatur zu Beginn des Programms sollte sicherstellen, dass genug Bakterien lysiert werden und ihre DNA freigesetzt wird. Die Annealing-Temperatur wurde den verwendeten Primern angepasst.

Die PCR-Produkte wurden mittels Agarosegelelektrophorese überprüft. Von jedem PCR-Ansatz werden 6 µl mit 6 µl Ficoll-Auftragspuffer gemischt und aufgetragen. Als Marker diente die 1kb Plus-Ladder von Invitrogen. Typische Laufbedingungen waren auf einem 1%igen Agarosegel in 1X TBE 40 min mit 150 V. Mit dem Dokumentationsgerät von Biorad wurde ein Foto gemacht und ausgewertet. Wenn positive Reaktionen gefunden wurden, wurde wie oben beschrieben fortgefahren, der Vorgang war der gleiche, nur dass in jeden PCR-Ansatz nun nur 1 Klon gepickt wurde.

3.5.4 Glycerinkulturen

Um einen bestimmten Bakterienstamm, insbesondere einen Fosmid-Klon, der im Screening gefunden wurde, zu konservieren, wurden Glycerinkulturen angelegt. Diese können tiefgefroren gelagert und bei Bedarf wieder aufgetaut werden.

Von den Bakterien wurde eine Übernachtskultur im passenden Flüssigmedium angelegt. Ein Aliquot dieser Kultur wurde dann in einem dafür vorgesehen Röhrchen mit Glycerin (Endkonzentration 25%) gemischt und bei -80°C gelagert.

Als Reserve wurden stets mehrere solcher Glycerinkulturen für einen Stamm angelegt.

3.5.5 Isolierung der Fosmide

Um die Fosmide sequenzieren zu können, war es zunächst erforderlich, sie aus den EPI300-Bakterien zu isolieren. Hierzu wurde ein hauseigenes Protokoll verwendet, welches auf den Grundlagen von Birnboim und Doly basiert (Birnboim & Doly, 1979). Für das Protokoll werden die drei Lösungen B&D I-III benötigt (siehe 2.3: Chemikalien und Materialien).

Zunächst wurde der entsprechende Klon über Nacht kultiviert. 20 ml LB-Medium wurden in einen 100 ml-Kolben gefüllt und mit 20 µl Chloramphenicol-Lösung versetzt (Endkonzentration 25 µg/ml). Der Klon wurde eingerührt und der Kolben über Nacht bei 37°C inkubiert.

Um die Ausbeute an Fosmid-DNA zu erhöhen, wurde nun die Multi-Copy-Induktion verwendet, wie sie vom Hersteller angeboten wird. Dadurch sollen die Bakterien ihr jeweiliges Fosmid stark vermehren.

500 µl Übernachtskultur wurden in einen frischen 50 ml-Kolben pipettiert, in dem 5 ml LB-Medium plus 5 µl Chloramphenicol-Lösung vorgelegt waren. Der Kolben wurde auf dem Schüttler mit 180 rpm bei 37°C für 5 Stunden bebrütet.

Nun wurden die Bakterien pelletiert. Je 1,5 ml wurden in ein Cap pipettiert, dann wurde für 10 min mit 10.000 rpm zentrifugiert. Der Überstand wurde verworfen, es wurden erneut 1,5 ml Übernachtskultur hinzugefügt, und erneut wurde für 10 min mit 10.000 rpm zentrifugiert, und der Überstand erneut verworfen. Am Boden des Caps war nun das Pellet mit den Bakterien zu sehen.

Nun wurden 0,8 ml TNE-Puffer hinzugegeben und das Pellet wurde mit der "Waschbrett-Methode" gelöst.

Es wurde erneut zentrifugiert, diesmal mit 5.000 rpm für 5 min bei 4°C. Der Überstand wurde sauber abpipettiert und verworfen.

Es wurden nun 100 µl B&D I hinzugegeben und das Pellet wurde erneut mit der "Waschbrett-Methode" gelöst. Die Lösung war nun milchig-wässrig.

Es folgte eine Inkubation bei 37°C für 10 min.

Es wurden nun 200 µl B&D II hinzugegeben und das Cap wurde 4 mal invertiert. Danach folgte eine Inkubation auf Eis für 2 min. Die Lösung wurde wieder durchsichtig.

Es wurden 150 µl B&D III hinzugegeben und das Cap wurde 4 mal invertiert. Die Lösung flockte nun aus.

Es folgte wieder eine Inkubation auf Eis für 10 min. Dann wurde mit 14.000 rpm für 15 min bei 4°C zentrifugiert.

Der Überstand wurde sauber abgenommen und in ein neues Cap gefüllt. Hierbei war wichtig, möglichst nichts von den ausgefällten Proteinen (weiße Flocken) mitzunehmen.

Es wurde zur Lösung das gleiche Volumen – in etwa 450 µl – 100%iges Ethanol gegeben. Die Lösung wurde für 25 min bei -20°C inkubiert.

Es wurde mit 14.000 rpm für 15 min bei 40°C zentrifugiert. Auf dem Boden des Caps war nun unscheinbar ein kleines Pellet zu sehen. Der Überstand wurde mit einer Pipette verworfen.

Das Pellet wurde 5 min in der beheizten Speedvac getrocknet.

Danach wurde es in 100 µl QIAgen-Puffer P1 gelöst, leichtes Schnipsen hob das Pellet vom Boden des Caps. Der QIAgen-Puffer P1 enthält RNase, welche die RNA abbauen soll, die später die Sequenzierung stören kann.

Die Lösung wurde kurz abzentrifugiert. Dann folgte eine Inkubation bei 37°C für 10 min.

Es wurden 300 µl H₂O zugefügt.

Es wurden 400 µl Phenol-Chloroform-Isoamylalkohol zugefügt. Die Mischung wurde 3 min lang invertiert.

Dann wurde mit 13.000 rpm für 5 min bei Raumtemperatur zentrifugiert. Der Überstand wurde in ein neues Cap gefüllt. Die Mixtur bestand zu diesem Zeitpunkt aus 2 Phasen. Die obere Phase wurde sauber abgenommen und in ein neues Cap gefüllt.

Es wurden 400 µl Chloroform zugegeben. Wieder wurde 3 min lang invertiert. Danach wurde wieder mit 13.000 rpm für 5 min bei Raumtemperatur zentrifugiert und der Überstand wieder in ein neues Cap pipettiert.

Es wurden 2 Volumina 100%iges Ethanol zugegeben sowie $\frac{1}{10}$ Volumen 3 M Na-Acetat. Zum Mischen wurde das Cap sanft geschnipst.

Die Lösung wurde für 30 min bei -80°C inkubiert.

Es wurde mit 14.000 rpm für 15 min bei 4°C zentrifugiert.

Der Überstand wurde mit der Pipette verworfen.

Das Pellet wurde in der beheizten Speedvac 5 min getrocknet.

Zuletzt wurden 50 µl H₂O zugegeben um die DNA zu lösen.

3.5.5.1 Überprüfung der Ausbeute

Das isolierte Fosmid wurde mit dem Nanodrop-Gerät gemessen und auf einem Agarosegel sichtbar gemacht. Zur Messung mit dem Nanodrop wurden 1,6 µl auf dem Gerät aufgetragen. Für die Gelelektrophorese wurde ein 0,7%iges Agarosegel in 1X TAE verwendet, welches 60 min mit 120 V lief. Aufgetragen wurden 1 µl Probe mit 1 µl Ficoll-Auftragspuffer.

3.6 Bioinformatische Verarbeitung

Die Aufgabe der Bioinformatik war es zunächst, die Sequenz-Rohdaten zu einem Genom zusammenzusetzen.

Aus dem 454-Sequencer kamen die Reads in Form von *.fastq-Dateien. Diese beinhalten die Basensequenzen mit assoziierten Qualitätswerten, die für die einzelnen Basen die Sicherheit des Basecalls angeben. Alternativ können die Reads auch als *.sff-Datei ausgegeben werden, in diesem binären Dateiformat sind die Reads als "Flowgram" enthalten. Im Unterschied zur *.fastq-Datei beinhaltet die *.sff-Datei gewissermaßen die Auslesung der Maschine, die einer Base zugrunde liegt, allerdings ist die Datei dadurch deutlich größer.

Die Sanger-Reads wurden als *.scf-Datei geliefert, dieses Format beinhaltet ebenfalls die Basenabfolge und das zugrundeliegende Flowgram (Dear & Staden 1992).

Die Basensequenzen mussten dann aufgrund von übereinstimmenden Basenabfolgen überlappend aneinandergelagert werden. Daraus resultierten so genannte Contigs, also Teilstücke des Gesamt-Genoms, zwischen denen Lücken mit bislang unbekannten Sequenzen lagen. Die Reihenfolge dieser Contigs zueinander war zunächst unklar. Im vorliegenden Fall lag ein verwandter Stamm mit bekanntem Genom vor. So konnten die Contigs mit diesem verglichen werden, und in eine diesem entsprechende Reihenfolge gebracht werden. Dies nennt man auch Mapping. Nun konnte gezielt begonnen werden, die Lücken zwischen den einzelnen Contigs mittels passender PCRs zu schließen.

Die Reads dieser PCRs wurden als *.ab1-Dateien generiert und nach und nach in das wachsende Pseudogenom integriert – auch hier wieder an der Stelle mit maximal möglicher identischer Überlappung.

Ziel war es, alle Contigs zu einem einzigen Contig zu verbinden: der kompletten DNA-Sequenz des zirkulären Chromosoms des Bakteriums.

Diese reine Abfolge von DNA-Basen sollte daraufhin nach Sequenzen durchsucht werden, welche Gene darstellen. Dazu wurde ein ORF-Caller verwendet, ein Programm, welches offene Leseraster für Proteinbiosynthese findet. Die gefundenen CDS (codierende Sequenzen) sollten dann mit verschiedenen Datenbanken verglichen werden, um ihnen mögliche Genprodukte zuweisen zu können (Annotation).

3.6.1 Newbler Assembler (Genome Sequencer FLX Software, Roche)

Das Programm Newbler (GS *De Novo* Assembler Version 1.1.02.15) wurde zur *de novo*-Assemblierung der Sequenzdaten der Plasmid-Shotgun-Bibliothek (Sanger-

Reads) und des 454-Sequenzierlaufs (454-Reads) benutzt. Es wurde von Roche 454 life sciences zur Verfügung gestellt, der gleichen Firma, welche auch die Sequenzierung mittels Pyrosequenzierung vornahm. Das Programm verbindet die vielen einzelnen Reads zu größeren Contigs, indem es homologe Reads übereinander schiebt – insgesamt eine sehr rechenintensive Aufgabe. Das Programm gibt als Ergebnis eine *.fna-Datei aus, in der die resultierenden Contigs im FASTA-Format enthalten sind: 454allcontigs.fna.

Zudem gibt es numerische Ergebnisse der Assemblierung aus, die Anzahl und durchschnittliche Länge der verwendeten Reads etwa.

Zur weiteren Verwendung wurden nur Contigs mit einer Länge von mindestens 500 bp verwendet. Newbler gibt hierfür ebenfalls eine *.fna-Datei mit nur diesen Contigs aus: 454largecontigs.fna.

3.6.2 Mapping der Contigs mit MAUVE

MAUVE ist ein Programm, welches mehrere Genome zum Vergleich aneinander ausrichten kann. Die Genome werden untereinander dargestellt. Homologe Sequenzen werden in der gleichen Farbe dargestellt und durch Linien verbunden. Sequenzen, die homolog sind, aber auf unterschiedlichen Strängen liegen, werden punktgespiegelt dargestellt. Der Grad an Übereinstimmung wird durch stärkere Ausfüllung des farbigen Balkens, der die Sequenz darstellt, angezeigt. Eine Sequenz, die nur in einem der Genome vorkommt, wird dementsprechend weiß angezeigt. (Darling et al. 2004)

Das Programm lässt sich dazu nutzen, die Contigs eines Genomprojektes an einem Referenzgenom auszurichten. Für den *E. faecalis* Symbioflor 1 wurde das Genom von *E. faecalis* V583 als Referenz verwendet. Dazu wurde die entsprechende *.gbk-Datei (*Enterococcus_faecalis_V583.gbk*) von der Webseite des NCBI heruntergeladen und in MAUVE als eines der zu vergleichenden Genome geöffnet. Als zweites Genom wurden die großen Contigs aus der Newbler-Assemblierung geladen (454largecontigs.fna). MAUVE wurde mit Standard-Parametern gestartet.

Das Ergebnis war eine graphische Darstellung der Contigs von *E. faecalis* Symbioflor 1 und dem Genom von *E. faecalis* V583. Diese diente dazu, die Contigs zu ordnen: Es wurde von Hand eine *.txt-Datei (mappedcontigs.txt) erstellt, in der die Contigs, die eine Entsprechung im Referenzgenom fanden, in der gleichen Reihenfolge wie im Referenzgenom aufgelistet wurden. Hinter jedes Contig wurde, per Tabulator-Taste getrennt, ein "+" oder ein "-" hinzugefügt. Das Pluszeichen bedeutete eine richtige

Orientierung im Sinne der Referenz, das Minus bedeutete, das Contig lag auf dem Gegenstrang und musste noch revers komplementiert werden. Diese so formatierte Datei benötigte das Programm minimap.

3.6.3 Primerwalking mit minimap

Das Programm minimap ist ein im Haus geschriebenes Programm für den Zweck, in dieser Phase eines Genomprojektes die Contigs zu ordnen und Primer für die nachfolgende Phase des Primerwalking zu generieren.

Das Programm erhielt als Input die oben erwähnte Datei mit allen großen Contigs (454largecontigs.fna) und die manuell mit Hilfe von MAUVE generierte Liste mit der Reihenfolge und Orientierung derselben (mappedcontigs.txt). Außerdem benötigte es eine Referenz um den OriC zu finden: Hierfür wurde die Datei *Enterococcus_faecalis_V583.fna* (das Genom von *E.faecalis* V583 im FASTA-Format, ebenfalls von der Webseite des NCBI) verwendet; dem Programm wird als Referenz der Anfang dieser Sequenz gegeben, dort liegt der OriC.

Das Programm liest die Contigs ein und ordnet sie nach der erstellten Liste in der richtigen Reihenfolge. Die mit Minus-Zeichen versehenen Contigs werden vom Programm revers komplementiert. Außerdem spaltet es das Contig, das den OriC enthält:

Dann sucht minimap Primerpaare für die Lückenschluss-PCRs.

Das Programm sucht dabei Primer, die mindestens 150bp vom Rand des Contigs entfernt liegen. Dieser Mindestabstand ist wichtig, da sonst die Reads unter Umständen keine Überlappung mehr mit dem Contig haben. Zum einen wird der Primer selbst nicht mit sequenziert, zum anderen sind meistens die ersten 20 bis 30 Basen eines Reads nicht eindeutig und sauber zu lesen. Aus diesen Gründen wäre es nicht sinnvoll, den Primer zu nahe an der Lücke zu wählen – obwohl man ihn natürlich möglichst nahe haben möchte, um weiter in die Lücke hinein sequenzieren zu können. Ein Abstand von 150 bp bis 450 bp ist ein Kompromiss der sich in der Praxis bewährt hat.

Minimap sucht innerhalb dieser Grenzen nach möglichen Primern. Diese werden dann gegen das komplette Pseudogenom geblastet.

Als Ergebnis wird folgendes ausgegeben:

- alle gemappten Contigs im FASTA-Format (mappedcontigs.fas)
- alle nicht gemappten Contigs im FASTA-Format (unmappedcontigs.fas)
- alle Contigs mit den gemappten am Anfang und den ungemappten am Schluss (all_contigs_sorted.fas)
- eine Liste von Primern für jede Lücke zwischen gemappten Contigs. Zu jedem Primer werden die Blast-Ergebnisse mit ausgegeben.

Dem Benutzer ist es nun überlassen, für jede Lücke zwei passende Primer auszuwählen. Ein idealer Primer bindet nur an der einen Stelle, für die er vorgesehen ist; die mitgelieferten Blast-Ergebnisse ermöglichen, schnell ungeeignete Primer aufgrund von mehrfachen Treffern im Genom auszusortieren. Bleibt für eine Seite einer Lücke kein geeigneter Primer, so kann das Programm mit erweiterten Grenzen für die Suche gestartet werden.

Für *E.faecalis* Symbioflor 1 erhielten die Primer Namen, die die Nummer der Lücke enthielten sowie eine fortlaufende Nummer. Sie wurden mitsamt Sequenz in eine Excel-Tabelle aufgenommen (Alle Primer.xls) und bei einem kommerziellen Anbieter (eurofins) bestellt.

Zur Kontrolle der richtigen Übernahme der Ausrichtung der Contigs – und damit der richtigen Zuordnung der Primerpaare – wurden die Contigs, wie sie von minimap anhand der Liste geordnet wurden, mit MAUVE erneut mit dem Referenzgenom verglichen. Die Datei all_contigs_sorted.fas wurde analog zum oben beschriebenen Prozess mit dem Genom von *E. faecalis* V583 verglichen. Der grafische Output zeigte nun die gemappten Contigs in der richtigen Reihenfolge, alle auf dem oberen Strang, sowie hintendran die nicht gemappten Contigs.

3.6.4 Assemblierung von PCR-Reads

Das Assembly wurde für den Lückenschluss in das Programm Seqman (DNASTar Lasergene 8) geladen.

Im Gegensatz zu Newbler bietet Seqman eine grafische Benutzeroberfläche für die weitere Bearbeitung der Contigs sowie einige nützliche Features wie zum Beispiel, halbautomatisch Primer für das Primerwalking zu entwerfen.

Hierzu wurden zunächst die Contigs aus der Newbler-Assemblierung in Seqman geladen: die Datei 454largecontigs.fna wurde ins Hauptfenster eines neuen Projektes gezogen. Es öffnete sich ein Fenster mit den Contigs. Damit diese nun nicht fälschlicherweise miteinander assembliert werden, wurden sie zunächst als Referenz markiert (Rechtsklick -> mark as reference). Dann wurden sie durch Klick auf „Assemble“ in das Projekt integriert. Sie erschienen im Fenster „Reads“. Dort wurden sie alle markiert und nicht mehr als Referenz markiert (Rechtsklick -> unmark reference).

Das Projekt wurde zur Sicherheit gespeichert.

Neue Reads wurden von der Firma AGOWA als *.ab1-Dateien zum Download bereitgestellt. Die Dateien wurde in ein mit Datum versehenes Unterverzeichnis gespeichert. Von dort konnten sie per Drag&Drop in das aktuelle Projekt eingefügt werden.

Dann musste jeder Read zunächst überprüft werden:

In Seqman wird der Read als Chromatogramm mit den zugehörigen Basecalls angezeigt. Undeutliche Chromatogramme, die zum Beispiel durch Überlagerungen durch doppelte PCR-Produkte oder verunreinigte Proben entstehen, zeigen schlechte Basecalls und damit schlechte Sequenzen an. Solche schlechten Reads wurden im Zweifelsfall verworfen.

Zudem wurden Sequenzen getrimmt: Der Anfang jeder Sequenz beinhaltet in der Regel zwischen 20 und 30 bp mehrdeutige, nicht verwertbare Basecalls, erkennbar am Chromatogramm, was an dieser Stelle Kurven zeigt, die nicht sauber einzelnen Basen zugeordnet werden können. Gegen Ende der Sequenz wird die Eindeutigkeit der Basecalls ebenfalls immer schwächer. Im Programm wurde per Hand eine Grenze zu Beginn und gegen Ende gesetzt, sodass nur diesseits der Grenzen liegende Sequenz von Seqman zur Assemblierung verwendet wurde.

Die auf diese Weise getrimmte Sequenz sollte nun in das Assembly integriert werden.

Zunächst wurden die Parameter von Seqman eingestellt. Verwendet werden kann entweder der Classic Assembler oder der Pro Assembler. Da schon getrimmt wurde, wurde die automatische Trim-Funktion von Seqman ausgeschaltet. Im Regelfall blieben die Parameter für die Assemblierung auf den Standardwerten und es wurde der Pro Assembler genutzt. In schwierigen Fällen wurde die zumindest benötigte

Übereinstimmungs-Prozentzahl oder Fragmentlänge variiert und das Assembly von Hand bearbeitet.

Im wesentlichen ging es an dieser Stelle darum, zu überprüfen, dass der Read an genau einer Stelle im Assembly passt und an keiner anderen. Zudem sollte er mit möglichst hoher Übereinstimmung passen. Im Zweifel wurde hier das Chromatogramm in Seqman betrachtet; vermeintliche SNPs konnten so zum Beispiel auf ein schlechtes Basecalling aufgrund eines an dieser Stelle unsauberes Chromatogramm zurückgeführt werden. In diesen Fällen wurde die Basen manuell editiert.

Zur Überprüfung wurde die zu assemblierende Sequenz noch mit dem Tool „Blastit“ (ein im Institut entwickeltes Programm, um lokale Blasts zu ermöglichen) gegen das komplette aktuelle Assembly geblastet. Idealerweise sollte nur ein guter Treffer vorhanden sein. An dieser Stelle konnte die Sequenz dann eingefügt werden.

Die Kontrolle war nötig, weil Seqman die Sequenz an der Stelle einfügt, die anhand der verwendeten Parameter vermeintlich am besten passt. Seqman begeht dabei aber hin und wieder Fehler, und setzt Sequenzen an Positionen, mit denen sie eine geringere Übereinstimmung haben als mit einer anderen, der vermutlich richtigen Stelle.

Zudem hat es sich herausgestellt, dass Seqman manchmal, vermutlich aufgrund großer Datenmengen oder eines Bugs, ganze Sequenzen von Contigs abschneidet – ohne entsprechende Fehlermeldung.

Es war daher ratsam, folgendermaßen vorzugehen. Wurde durch einen ersten Assemblierungsversuch oder spätestens durch die Bestätigung mit Blastit festgestellt, dass eine Sequenz einwandfrei an eine bestimmte Stelle gehörte, wurde das aktuelle Seqman-Projekt geöffnet und die entsprechende Sequenz hinzugefügt. Dann wurden alle Contigs gesperrt, die nichts mit der aktuellen Sequenz zu tun hatten (Klick auf das kleine Symbol neben dem Contig, es wird zu einem Schloss), und erst dann wurde assembliert. So wurde Seqman daran gehindert, sich mit den falschen Contigs zu befassen. Dies sollte natürlich nur dann getan werden, wenn deutlich war, dass der Read wirklich nicht zu den anderen Contigs gehörte – daher der Blast gegen das Assembly.

Nachdem die Reads so assembliert wurden, wurde das Projekt unter einer fortlaufenden Nummer und mit dem aktuellen Datum versehen gespeichert. Zum Zwischenvergleich wurden die aktuellen Contigs aus Seqman als *.fna-Dateien

exportiert und in Mauve mit der Referenz verglichen. So konnte der Prozess überprüft und korrigiert werden.

3.6.5 Primerwalking mittels Seqman

Nachdem der erste Schwung Primer mittels des Programms minimap ausgegeben wurde, wurde für die Fortsetzung des Primerwalkings die Funktion in Seqman verwendet. Nach dem oben beschriebenen Schritt der Assemblierung einer neuen Sequenz war das entsprechende Contig normalerweise ein Stück länger, bzw. die Lücke ein wenig kleiner.

Unter dem Menüpunkt Contig -> Primer Walk wurden von Seqman Primer generiert. Hier wurden die Standardeinstellungen gewählt mit einer Ausnahme: „Avoid locations from end“ wurde auf mindestens 150 hochgesetzt – analog zu der genannten Einstellung bei minimap.

Seqman gab nun für beide Enden des Contigs Primer aus. Der für die entsprechende Lücke bestimmten Primer wurde nun noch einmal mit Blastit geprüft, indem er gegen das momentan aktuelle Assembly geblastet wurde. Idealerweise sollte der Primer nur ein einziges Mal binden, nämlich am Rande der entsprechenden Lücke. Gab es Homologien zu mehreren Stellen, wurde ein neuer Primer gesucht. Dazu konnte zunächst wieder Seqman verwendet werden, indem man die Entfernung zur Lücke weiter erhöhte, sodass das Programm weiter entfernte Primer suchte. Hatte dies weiter keinen Erfolg, wurde auf Primer 3 ausgewichen.

3.6.6 Primerwalking mittels Primer3Plus

War mit Seqman kein Primer zu finden, der zu genau einer Stelle homolog war, wurde auf die Software Primer3Plus zurückgegriffen. Diese ist im Internet zu finden unter: <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>

Der Rand des Contigs – etwa 1.000 bp - wurde mit Kopieren & Einfügen aus dem Contig in Seqman in das „source sequence“-Eingabefeld eingefügt.

Als „Task“ wurde „Primer List“ angegeben. An den sonstigen Einstellungen war zunächst keine Änderung vorzunehmen. Bei Bedarf wurde zusätzlich eingestellt, welche Regionen innerhalb der Zielsequenz nicht für die Suche verwendet werden sollten. Für die Einstellung "left" und "right" wurde nur die Option angewählt, die den Primer in Richtung des Gaps lieferte.

Das Programm gab eine Liste von mehreren Primern aus. Diese wurden mittels Blastit getestet. Auch hier galt: Ein optimaler Primer sollte genau einen guten Treffer an der richtigen Stelle im Genom haben.

An manchen Stellen musste auch bei der Primersuche mit Primer3Plus die Entfernung zur Lücke verlängert werden, indem dem Programm eine längere Sequenz des Contig-Rands zur Verfügung gestellt wurde.

Die Primer wurden in einer Excel-Tabelle gespeichert und bestellt.

Für die Benennung der Primer galt folgendes Schema:

EF1_2.*n*_gp#_(*L/R*)_(*sm/p+*)

Hierbei stand *n* steht für eine fortlaufende Nummer, die die Phase des Primerwalking-Prozesses anzeigte. Das # stand für die Nummer der Lücke. *L* oder *R* stand dafür, auf welcher Seite der Lücke der Primer band. Die Kürzel "*sm*" oder "*p+*" standen dafür, mit welchem Programm der Primer ausgesucht wurde.

Beispiel: EF1_2.9_gp17_L_sm

In die Tabelle wurde der Name des Primers, dessen Basenfolge, und die Schmelztemperatur aufgenommen. Die Schmelztemperatur wurde mit Primer3+ berechnet (Option Primer_Check).

Anhand der Primertabelle konnte die Annealing-Temperaturen für die PCRs berechnet werden. Als Ausgangspunkt wurde sie um etwa 3°C niedriger angesetzt als die niedrigere Schmelztemperatur der beiden PCR-Primer.

3.6.7 Annotation des Genoms

Die DNA-Sequenz eines Organismus an sich ist eine bloße Abfolge von Basen. Um weitere Aussagen über das Genom treffen zu können, muss diese Basensequenz daher noch auf Gene untersucht werden. Hierzu werden zunächst Programme verwendet, die offene Leseraster in der DNA-Sequenz finden (englisch open reading frame, ORF). Diese sind prinzipiell als Grundlage für ein kodiertes Protein zu verstehen - wenn auch nicht alle dieser ORFs in der Realität zu einem fertigen Protein führen. Die DNA-Sequenzen dieser ORFs werden nach dem allgemeinen genetischen Code in entsprechende Abfolgen von Aminosäuren umgeschrieben. Diese Aminosäure-Sequenzen werden dann mit Datenbanken verglichen, in denen bekannte Proteine von den verschiedensten Organismen enthalten sind. Die Gene

werden dann anhand der besten Treffer in den Datenbanken annotiert, das heißt, man trifft aufgrund von Sequenz-Homologien mit anderen Organismen eine Vorhersage, welche Art von Protein dieses Stück DNA-Sequenz codiert. Dementsprechend gilt diese Gen-Beschreibung auch nur eingeschränkt: Es ist kein eigentlicher Protein-Nachweis, denn über die Expression des Proteins wird keine Aussage getroffen.

Im Falle von *E. faecalis* Symbioflor 1 konnte auch hier der Referenzstamm *E. faecalis* V583 genutzt werden. Sequenzen, die mit der Referenz überwiegend übereinstimmten, konnten entsprechend annotiert werden.

Für das Genom von *E. faecalis* Symbioflor 1 wurde folgendes halbautomatisches Vorgehen verwendet:

Die Contigs des aktuellsten Assembly wurden aus Seqman exportiert als *.fas-Datei. Diese wurde in GenDB hochgeladen und von diesem Programm automatisch annotiert (F. Meyer 2003).

Das Programm arbeitet mit einer Kombination aus GLIMMER, CRITICA und tRNAscan um codierende Regionen zu finden (Delcher 1999; Badger & G. J. Olsen 1999; Lowe & S R Eddy 1997). Die Software vergleicht diese codierenden Sequenzen anhand verschiedener Datenbanken wie zum Beispiel Swiss-Prot, KEGG, COG, Interpro und Pfam (Amos Bairoch et al. 2004; Tanabe & Kanehisa 2012; Tatusov 2000; Apweiler et al. 2001; Punta et al. 2012). GenDB annotiert das Genom zunächst automatisch anhand der besten Treffer.

Die vorläufige Annotation des Genoms wurde aus GenDB in Tabellenform exportiert.

Nun erfolgte der Abgleich beider Genome mit dem Programm GECO. Das Programm dient zum visualisierten Vergleich der Genome verschiedener Organismen auf Gen-Ebene in Bezug auf deren Aminosäure-Sequenz. Es sucht für jedes Gen jedes eingegeben Organismus homologe Gene in den Vergleichsgenomen und zeigt diese untereinander an. Diese Homologie-Zuordnung kann auch in Form einer Tabelle ausgegeben werden: In der Tabelle ist für jedes Gen eines Organismus angegeben, welche Gene des anderen Organismus zu diesem homolog sind. Hierbei können verschiedene Grenzwerte gesetzt werden für die Abdeckung und Identität der zu vergleichenden Gen-Sequenzen (C. T. Kuenne et al. 2007).

Auf diese Weise wurde zunächst eine Liste generiert, die alle Gene von *E. faecalis* Symbioflor 1 und *E. faecalis* V583 zeigte, die mindestens 60% Identität und mindestens 80% Abdeckung (bezüglich Aminosäure-Sequenz) zueinander aufzeigten. Von dieser Liste wurden all jene nicht berücksichtigt, die Paralogie aufweisen, das heißt, von denen in einem der Genome mehr als ein Treffer gefunden wurde.

Die übrige Liste der Gene, die genau eine Entsprechung im jeweils anderen Genom haben, wurde nun von Hand überprüft: Stimmte die automatische Annotation von GenDB mit der Annotation des homologen Gens von *E. faecalis* V583 überein, dann wurde in der Tabelle markiert, dass die Annotation übernommen wird.

Die Liste mit diesen Genen und deren Annotation wurde in GenDB hochgeladen und vom Programm übernommen.

Der Schritt wurde noch einmal wiederholt mit 40% Identität und mindestens 80% Abdeckung, wobei aus dieser Liste alle Gene herausgenommen wurden, die schon beim obigen Schritt annotiert wurden. Hiermit wurden die Gene annotiert, die ein Homolog in *E. faecalis* V583 aufzeigen, aber mit niedrigerer Übereinstimmung. Die Liste dieser Gene mit Annotation wurde ebenfalls in GenDB hochgeladen.

Die durch diesen Abgleich annotierten Gene erhielten in GenDB den Status: Regional status 3; functional status 1. Im Programm wurden sie rosa dargestellt.

Die übrigen Gene wurden in GenDB manuell überprüft:

Sie wurden zuvor wie genannt von GenDB automatisch annotiert und waren daher im Programm gelb-braun dargestellt und hatten den Status: regional 3, functional "automatically annotated".

Für jedes Gen wurden die von GenDB gefundenen Datenbanktreffer betrachtet und anhand der Qualität der Treffer sowie gegebenenfalls zusätzlich mit Blast-Vergleichen der Sequenz in der NCBI-Sequenz-Datenbank (<http://blast.ncbi.nlm.nih.gov>; Protein-Blast in der nucleotide collection) wurde das Gen annotiert. Es erhielt dann den Status: regional 3, functional "annotated" und wurde grün angezeigt.

Folgende Regeln wurden von einem früheren Genomsequenzierungs-Projekt Im Institut übernommen und für die manuelle Annotation verwendet (Steinweg 2005).

Bestimmte Funktion, bestimmte Spezifität und sichere Zuordnung

- Identity: >40%, Coverage: ganze Länge

Redundante Matches auf ganzer Länge zu einem bekannten Protein und einer Proteinfamilie: Hier wurde die Funktion, die EC-Nummer und der Genname übernommen.

- Coverage: teilweise

Redundante Matches mehrerer Regionen zu bekannten Proteinen und HMMs: Hier wurden die Funktionen, die EC-Nummern und der Genname übernommen und mit „/“ getrennt.

Bestimmte Funktion, bestimmte Spezifität und unsichere Zuordnung

- Identity: 25-40%

Schwacher Match zu einem bekannten Protein und HMM:

Hier wurde die Funktion übernommen und „putative“ angehängt.

Unbestimmte Spezifität (Familie, Gruppe)

- Identity: >40%

A) Match zu Mitgliedern einer Proteinfamilie mit bekannter Funktion: Hier wurde die Funktion übernommen und „Familie“ angehängt.

B) Match zu Mitgliedern einer Proteinfamilie mit unbekannter Funktion: Hier wurde die Familie übernommen und „protein“ angehängt.

C) Match zu einer Gruppe von Proteinen mit ähnlicher Funktion ohne Familie: Hier lag nur eine allgemeine Funktionsbeschreibung vor.

Unbestimmte Funktion (hypothetical protein)

- Identity: >40%

A) Match zu einer Domäne: Hier lag eine Domänenfunktion oder ein Domänenname vor

B) Coverage: ganze Länge: Conserved hypothetical protein

C) Coverage: teilweise: conserved domain protein

D) Drei oder mehr TMHMM-Regionen: membrane protein, putative

E) Match zu Lipoprotein motif: Lipoprotein, putative

F) Drei oder mehr TMHMM-Regionen + Lipoprotein motif: membrane protein, putative

G) Match zu signal-P protein: putative secreted protein

Keine Zuordnung möglich

- Identity: <25%: hypothetical protein

Für die Sequenzen nahe der Lücken, die später hinzukamen, wurde zusätzlich der ORF-Caller RAST zur Annotation verwendet (Aziz et al. 2008).

4 Ergebnisse

4.1 Assemblierung des Genoms von *E. faecalis* Symbioflor 1

4.1.1 454-Sequenzierung

Von der Firma Roche wurde genomische DNA von *E. faecalis* Symbioflor 1 mittels 454- Pyrosequenzierung sequenziert. Hierbei wurden 208.056 Reads generiert, mit insgesamt 51.900.784 Basen.

4.1.2 Assemblierung mit Newbler

Mit dem Programm Newbler (GS *de novo* Assembler) wurden seitens Roche die 454- Reads zusammen mit den Sanger-Reads aus der Plasmid-Bibliothek in einem Hybrid-Assembly assembliert.

Von den Sanger-Reads wurden 13.131 von 13.223 Reads (99,30%) assembliert.

Von den 454-reads wurden 206.726 von 208.056 Reads (99,36%) assembliert.

Insgesamt wurden demnach 219.857 von 221.279 Reads (99,36%) assembliert.

Es resultierten 130 Contigs mit insgesamt 2.773.036 bp, davon waren 69 Contigs mindestens 500 bp lang, diese waren insgesamt 2.758.197 bp lang. Die Basen der Consensus-Sequenz hatten zu 99,63% einen Qualitätswert von mindestens Q40 (maximal ein Fehler auf 10 kbp), die restlichen 0,37% hatten einen Qualitätswert von Q39 und niedriger (Fehlerrate höher als ein Fehler auf 10 kbp).

Die durchschnittliche Abdeckung betrug 20,95 (das heißt, im Schnitt wurde jedes Basenpaar 20,95-fach sequenziert).

4.1.3 Mapping mit MAUVE

Die Ausrichtung der Contigs am Referenzstamm *E. faecalis* V583 mit MAUVE ergab, dass 45 der 69 großen Contigs gemappt werden konnten, es verbleiben dazwischen 45 Lücken (Gaps).

4.1.4 Primerwalking mit minimap

Mit minimap wurden 45 Primerpaare für das Primerwalking entworfen (siehe Anhang). Mit diesen PCRs konnten zunächst 17 Lücken geschlossen werden, es waren nun 28 gemappte Contigs übrig.

4.1.5 Primerwalking mit Seqman und Primer3Plus

Es wurden insgesamt weitere 95 Primer entworfen (siehe Anhang).

4.1.6 Lückenschluss mittels PCR

Mit den Primern aus minimap, Seqman und Primer3plus konnten insgesamt 199 für den Lückenschluss verwertbare Reads generiert werden. Mit ihnen konnten 39 der 45 Lücken zwischen den gemappten Contigs geschlossen werden.

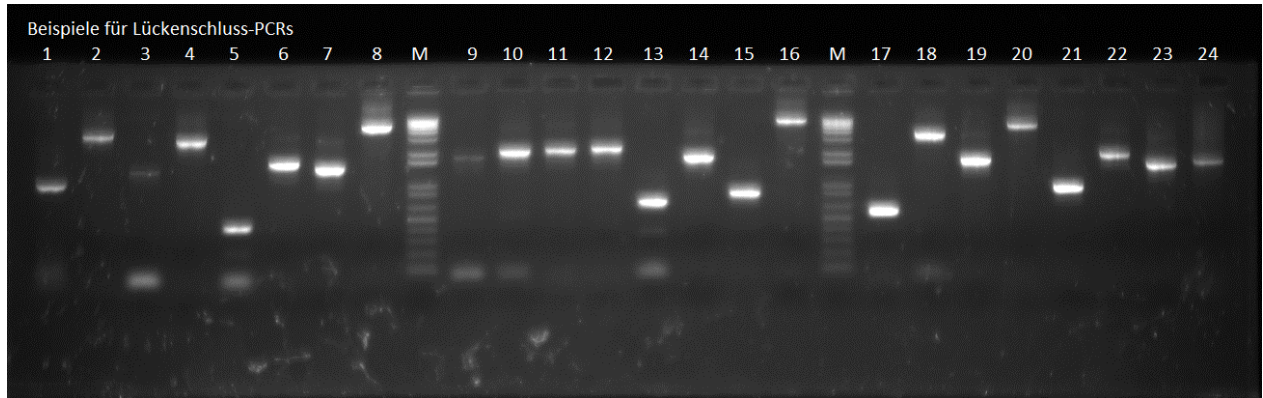


Abbildung 8: Lückenschluss-PCRs

Beispiel für PCRs für den Lückenschluss. Größere Produkte wie 2,4 und 8 wurden sukzessive mittels Primerwalking sequenziert. Marker: 1kb Plus

4.1.7 Sequenzierung der rRNA-Operons

Indem das wachsende Assembly mit MAUVE mit dem Referenzstamm verglichen wurde, stellte sich heraus, dass 4 der verbliebenen Lücken den 4 rRNA-Operons von *E.faecalis* V583 entsprachen. Da Bakterien ribosomale RNA zur Translation und damit zur Proteinsynthese benötigen, sind die entsprechenden Gene stark konserviert (Hillis & Dixon 1991). Daher war zu erwarten, dass auch *E. faecalis* Symbioflor 1 an diesen Stellen entsprechende rRNA-Operons besaß.

Die rRNA-Operons zeichnen sich durch hohe Homologien zueinander und interne Wiederholungsstrukturen aus. Dies erschwert dem Assembler bei der *de novo* Assemblierung die Zuordnung der Reads zum jeweils richtigen Operon. Auch in diesem Fall sind zunächst alle 4 Operons offene Lücken geblieben.

Um die rRNA-Operons zu sequenzieren, wurden deshalb für die PCRs Primer gewählt, die weit genug von den repetitiven Sequenzen der rRNA-Operons entfernt waren, um die Region jedes Operons gezielt einzeln amplifizieren zu können. Diese

PCR-Produkte wurden mit Standard-rRNA-Primern sequenziert, die im Institut für frühere prokaryontisches Genomsequenzierungen verwendet wurden (siehe Anhang: rRNA-Kassetten-Primer) (Steinweg, 2005).

Um diese Reads eines bestimmten Operons an die passende Stelle zu assemblieren, wurden in Seqman nur die Contigs am Rand dieses einen Operons freigegeben, alle anderen Contigs, insbesondere die am Rande der anderen Operons, wurden gesperrt. Dann wurden die Reads der dazu passenden PCRs geladen und assembliert. So wurde für jedes rRNA-Operon einzeln verfahren.

Auf diese Weise wurden 83 Reads generiert, mit denen die 4 Lücken mit den darin enthaltenen rRNA-Operons sequenziert werden konnten.

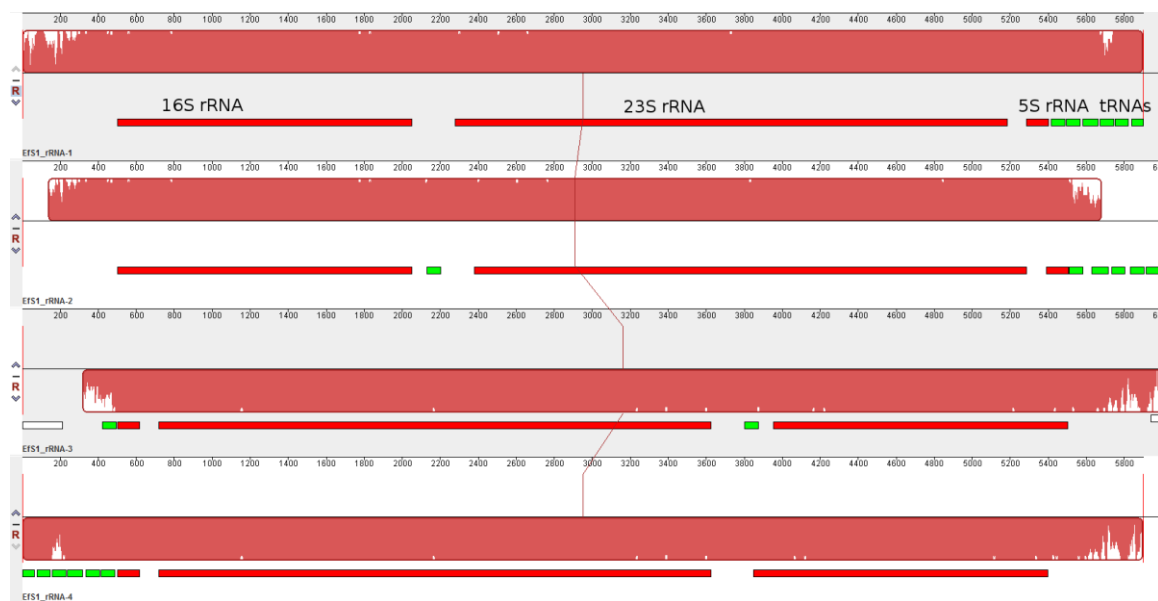


Abbildung 9: Die 4 rRNA-Operons des *E. faecalis* Symbioflor 1 in MAUVE

Jedes Operon besitzt Gene für eine 5S-rRNA, eine 16S-rRNA und eine 23S-rRNA. Die Loci sind sehr homolog zueinander. Zusätzlich vorhanden sind unterschiedliche tRNAs, in grün gezeigt. Locus 3 und 4 sind auf dem Gegenstrang angeordnet.

4.1.8 Fosmid-Bibliothek von *E. faecalis* Symbioflor 1

4.1.8.1 Herstellung der Fosmid-Bibliothek

Die Herstellung geschah nach dem beschriebenen Protokoll. Es wurden insgesamt 1.440 Klone für 15 96-well-Genbank-Platten gepickt. Von jeder Platte wurde eine Sicherheitskopie erstellt und alle Platten wurden bei -80°C aufbewahrt.

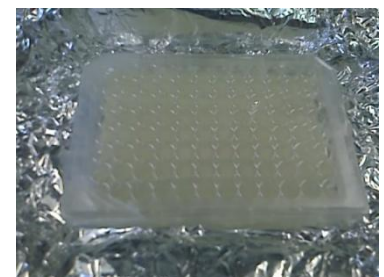


Abbildung 10: Fosmid-Genbank-Platte

Die Platten wurden mit Parafilm verschlossen, mit Alufolie eingewickelt und bei -80°C gelagert

Zur Kontrolle wurden stichprobenweise Fosmide isoliert und mit BamH1 verdaut. In der Agarose-Gelelektrophorese zeigten sich Inserts verschiedener Größen und mit unterschiedlichem Restriktionsverdau-Muster.

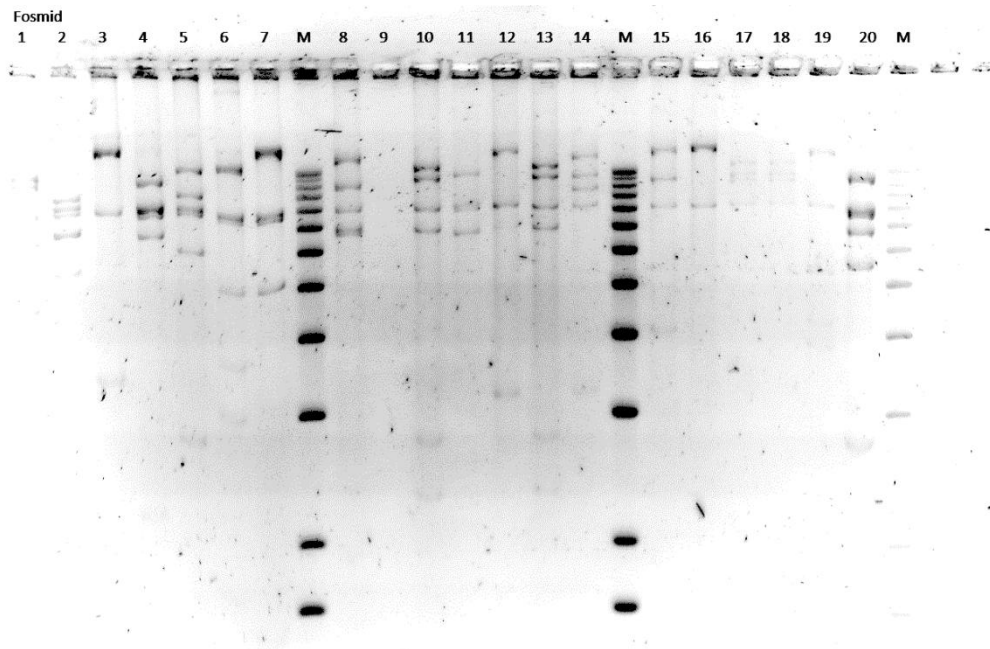


Abbildung 11: Fosmid-Restriktionsverdau mit BamH1

Es zeigt sich, dass manche Fosmide (hier 10 und 16 und 19 sowie 17 und 18) das gleiche Restriktionsmuster aufweisen und demnach vermutlich gleiche Klone sind. Marker: 1kb Plus

4.1.8.2 Screening der Fosmid-Klone

Es wurde nach Fosmid-Klonen gesucht, mit deren Inserts die Lücken zwischen verbliebenen Contigs geschlossen werden könnten.

Hierfür wurden für jede Lücke zwei Primerpaare entwickelt: Von jedem Contig wurden 1.000 bp von dem der Lücke zugewandten Rand abgeschnitten und in Primer3plus geladen. Neben den Standardeinstellungen wurde gewählt: "Detection", PCR-Produktgröße 300 bp.

Über BLAST-Vergleich mit dem aktuellen Assembly wurden unter den gefunden Primern die spezifischsten ausgesucht. Mit diesen wurden die Fosmid-Klone nach oben genanntem Protokoll gescreent.

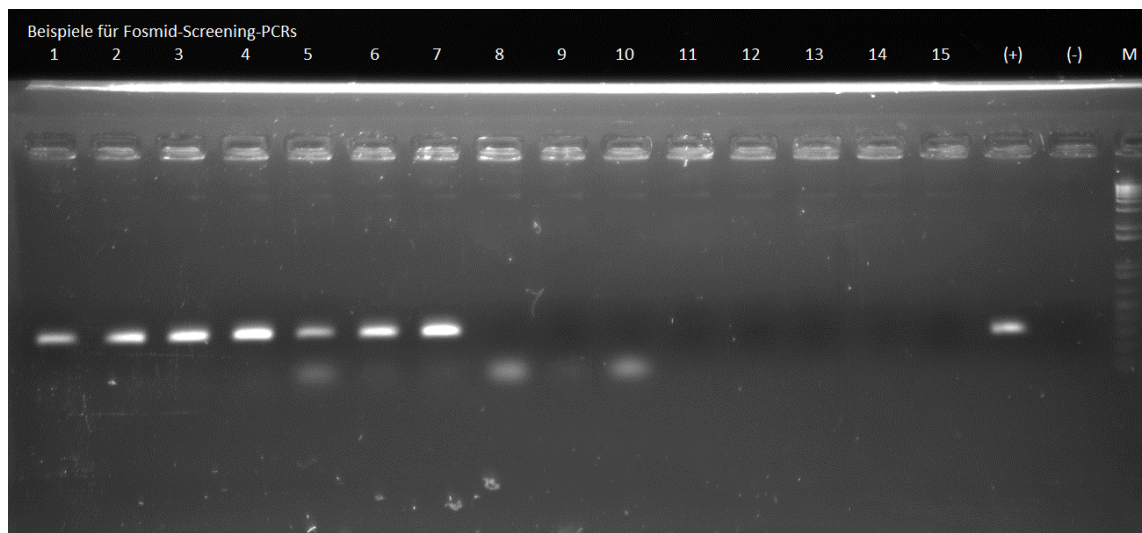


Abbildung 12: Fosmid-Screening-PCRs

Die positiven PCRs (1-7) zeigen an, dass die Klone ein Insert mit der gesuchten Sequenz enthalten. Marker: 1kb Plus

Fosmid	Contig 6 Ende	Contig 7 Anfang	Contig 91A	Contig 8 Anfang	Contig 8 Ende	Contig 21 Anfang	Contig 102A	Contig 20 Ende
1KA.D3	+	+			+			
2KB.A8	+	+	+					
2KB.A10	+	+						
2KB.C12	+	+	+					
2KB.F8	+	+						
2KB.F12	+	+			+			
2KB.G10		+	+					
1KB.D12				+				
1KB.F10				+	+			
1KB.F11				+	+			
2KB.D8				+	+			
2KB.E10				+				
2KB.B11							+	+

Tabelle 1: Ergebnis des Fosmid-Screenings

Das Pluszeichen steht für eine positive Screening-PCR

Die Fosmide wurden isoliert und mit Primern ansequenziert, die auf dem Vektor liegen.

Die Reads wurde in Seqman geladen und mit den aktuellen Contigs assembliert.

Aus jedem Fosmid wurden so 2 Reads gewonnen, die gegenläufig an den Enden des jeweiligen Inserts lagen. Passte jeder dieser beiden Reads zu einem unterschiedlichen Contig, bedeutete das, dass diese Contigs nebeneinander liegen mussten. Zudem konnte somit die Ausrichtung der Contigs bezüglich Strang und Gegenstrang bestimmt werden.

Mit diesen Informationen konnten neue Primerpaare zur Verbindung von Contigs gefunden werden und entsprechende PCRs durchgeführt werden:

Die Lücke zwischen Contig 20 und Contig 102A konnte auf diese Weise geschlossen werden. Ebenso konnte so Contig 7 an Contig 91A gehängt werden.

Die Lücke zwischen Contig 6 und Contig 7 ist zur Zeit noch offen. Die Sequenzierung der Fosmid-Inserts mit dem Vektor-seitigen Primerpaar lässt den Schluss zu, dass die Anordnung zueinander stimmt. Das Primerwalking in die Lücke hinein scheiterte jedoch bisher – ab einem gewissen Punkt wurden keine verwertbaren Reads mehr generiert.

4.1.9 Finaler Stand der Assemblierung

Bei Fertigstellung dieser Arbeit ist die Genomsequenzierung bei folgendem Stand angelangt:

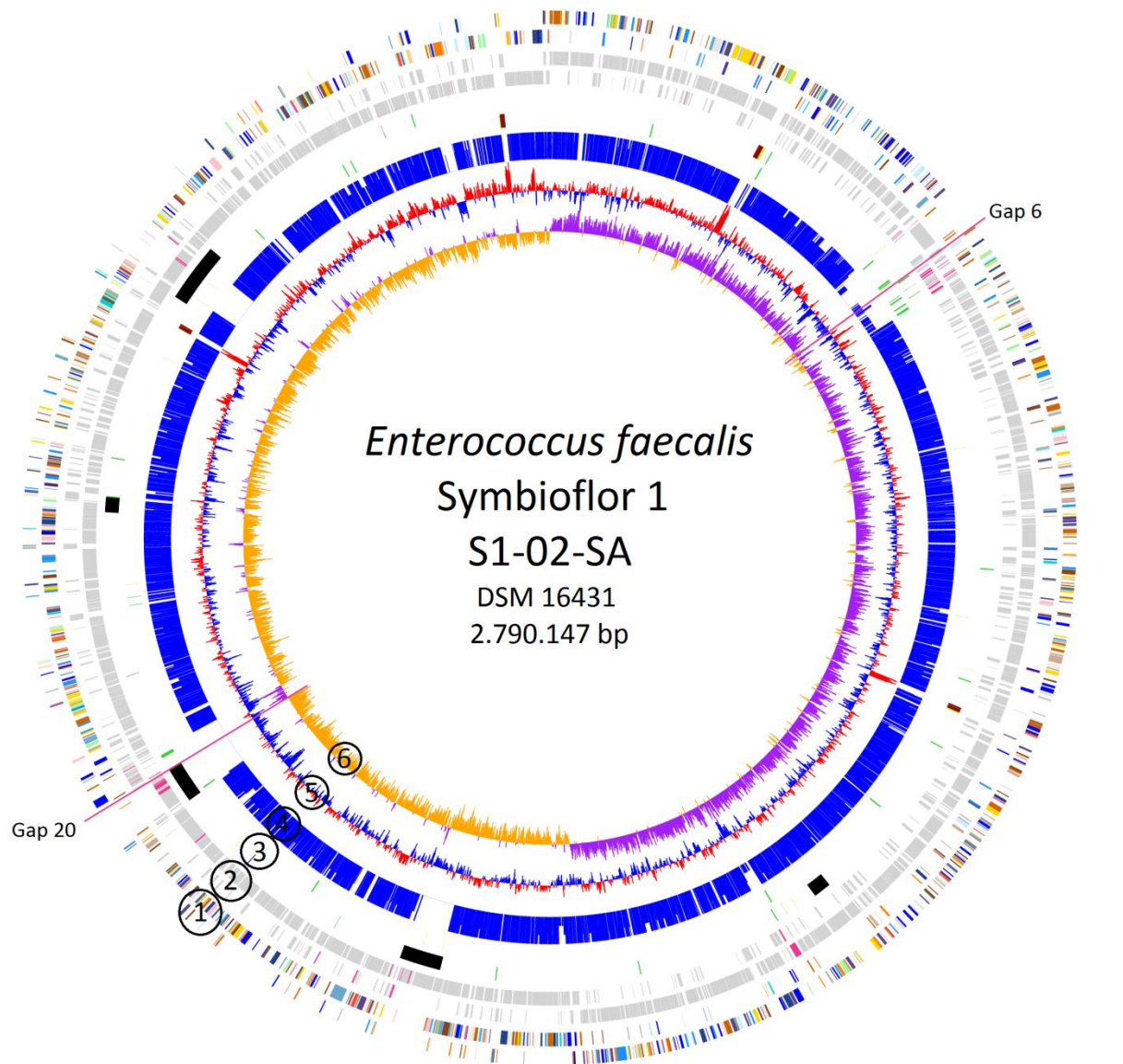
Name des Contig	Länge in bp
Contig 'uc084'	8.016
'Contig07091011121314151617181920uc102'	1.434.223
'Contig08A212223242526272829010203040506'	1.355.724
Total	2.797.963

Das Contig "uc084" ist eines der ungemappten Contigs und ist daher, auch aufgrund seiner eher geringen Größe, für die Annotation und weitere Analyse nicht verwendet worden. Die beiden großen Contigs wurden mit je 100 N als Platzhalter verbunden.

4.2 Allgemeine Charakteristika des Genoms von *E. faecalis*

Symbioflor 1

Das zirkuläre Chromosom mit 2 Lücken hat zur Zeit eine Länge von 2.790.147 bp. Der Stamm besitzt kein Plasmid. Der GC-Gehalt beträgt 37,7%. Es sind insgesamt 2.714 codierende Sequenzen detektiert worden (CDS). Diese haben einen GC-Gehalt von 38,2% und eine durchschnittliche Länge von 302 Aminosäuren. Damit sind 88,3% der DNA-Sequenz codierend. Es sind 12 rRNA-Gene enthalten, die in 4 Operons liegen. Daneben sind 63 tRNA-Gene vorhanden.



Von außen nach innen:

Ring 1: Gene von Symbioflor 1 nach COG-Klassen

Ring 2: Potentiell horizontal transferierte Gene in pink

Ring 3: Prophagen in schwarz, CRISPR in blau, Transposasen in grün, tRNA in grüngelb, rRNA in braun

Ring 4: Vorhandensein der Gene innerhalb der 3 Referenzstämme (60% Identität, 80% Abdeckung der AS-Sequenz)

Ring 5: Abweichung des GC-Gehalts

Ring 6: GC-Deviation (G-C/G+C)

Abbildung 13: Übersicht über das Genom von *E. faecalis* Symbioflor 1

Erstellt in GenomeViz (Ghai et al. 2004) und Augur (André Billion et al. 2006)

Folgende Tabelle gibt einen Überblick über allgemeine Charakteristika von *E. faecalis* Symbioflor 1 und den bisher drei vollständig sequenzierten *E. faecalis*-Stämmen.

<i>E. faecalis</i>	Symbioflor 1	V583	OG1RF	62
Chromosomgröße [bp]	2.790.147	3.218.031	2.739.625	2.988.673
G+C-Gehalt [%]	37,70	37,50	37,80	37,50
G+C-Gehalt in CDS [%]	38,20	38,10	38,20	38,10
Anzahl an CDS	2.716	3.112	2.579	2.897
Durchschnittliche Anzahl an Aminosäuren in einer CDS	302	296	316	299
Codierender Anteil [%]	88,30	86,20	89,50	87,20
Anzahl rRNA-Gene	12	12	12	12
Anzahl rRNA-Operons	4	4	4	4
Anzahl tRNA-Gene	63	68	58	55
Prophagen	5	8	1	4
davon intakt	2	2		1
davon inkomplett	2	6	1	2
davon fraglich	1			1
Plasmide	0	3	0	3
CRISPR Loci	3	1	3	2
davon bestätigt	1		2	
davon fraglich	2	1	1	2

Tabelle 2: Numerische Übersicht über die Genome der 4 Enterokokken-Stämme

Statistischer Output aus GeCo (C. T. Kuenne et al. 2007), Phagen-Analyse durch PHAST (Zhou et al. 2011), CRISPR Loci durch CRISPRFinder (Grissa et al. 2007)

4.3 Multi-Locus-Sequence-Typing

Eine gebräuchliche molekularbiologische Technik, Bakterien zu klassifizieren, ist das Multi-Locus-Sequence-Typing (MLST). Dabei werden, ähnlich wie der Typisierung anhand der rRNA-Gene, zunächst Gene bestimmt, die im Kerngenom einer Bakterienspezies enthalten sind. Für den *Enterococcus faecalis* wurden hierfür 7 Haushaltsgene gewählt, die vermutlich auch in neuen, unbekannten Isolaten enthalten sein werden. In diesen Genen wurde nach Loci gesucht, die so stark konserviert sind, dass Primer entwickelt werden konnten, die bei allen zu typisierenden Stämmen ein PCR-Amplifikat erzeugen können. PCR-Amplifikate eines Locus von verschiedenen Stämmen bilden Allele, die sich meist nur durch einzelne SNPs unterscheiden.

Das Profil der Allele der 7 Loci eines Stammes ergibt dessen MLST-Sequenztyp (Ruiz-Garbajosa et al. 2006).

Auf <http://efaecalis.mlst.net/> steht eine Datenbank zur Verfügung, in der solche MLST-Loci vorhanden sind und mit der eigene Stämme typisiert werden können. Dies wurde für die 4 untersuchten Stämme getan, indem deren Haushaltsgene aus GeCo exportiert wurden. Aus den Genen wurden die entsprechenden Loci mit MEGA5 unter Zuhilfenahme von Referenz-Allelen ausgeschnitten und dann online analysiert.

	<i>E. faecalis</i>			
Gen mit Allel des Stammes	Symbioflor 1	V583	OG1RF	62
<i>gdh</i>	2	12	3	3
<i>gyd</i>	7	7	1	3
<i>pstS</i>	11	3	16	9
<i>gki</i>	1	7	1	1
<i>aroE</i>	3	6	1	9
<i>xpt</i>	4	1	1	19
<i>yqiL</i>	2	5	1	7
Sequenz Typ	248	6	1	66

Tabelle 3: MLST der 4 Enterokokken-Stämme

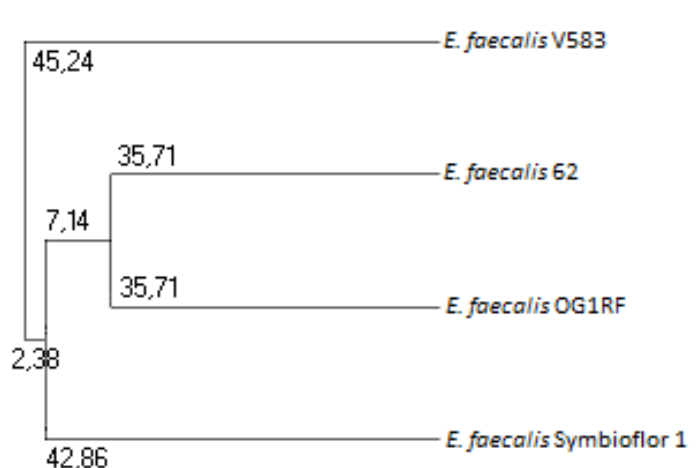


Abbildung 14: Phylogenetischer Baum anhand von MLST

4.4 Phylogenetische Untersuchung

Das MLST kann auch verwendet werden, wenn für einen Stamm nur PCRs für die 7 Loci vorhanden sind. Wenn komplette Genome vorhanden sind, kann die Analyse erweitert werden um mehr Sequenzdaten mit einzubeziehen.

Anstelle von 7 Loci wurde für die folgende Untersuchung das komplette Kerngenom der 4 Stämme herangezogen.

Zunächst wurde mit GeCo eine Homologie-Matrix der Gene aller 4 Genome erstellt (60% Identität der Aminosäuren, 80% Abdeckung). Alle Gene, die in allen Organismen genau ein Homolog besitzen, wurde für die weitere Analyse verwendet: Damit sind alle Gene erfasst, die in allen 4 Genomen jeweils genau einmal vorhanden sind. Die Sequenzen all dieser Gene eines Stammes wurden in einer *.fas-Datei zusammengefasst. Diese wurden mit Mugsy aneinander ausgerichtet (Angiuoli & Salzberg 2011). Die ausgerichteten Sequenzen wurden mit MEGA5 einer phylogenetischen Analyse unterzogen: Mit der neighbour-joining-Methode von 100 Bootstrap-Proben wurde ein phylogenetischer Baum erzeugt (Saitou & M Nei 1987; Felsenstein 1985; Tamura et al. 2004; Tamura et al. 2011). Zum Legen der Wurzel diente eine Wiederholung der Analyse unter Einbeziehung von *Lactobacillus reuteri* DSM 20016 (Copeland et al. 2007).

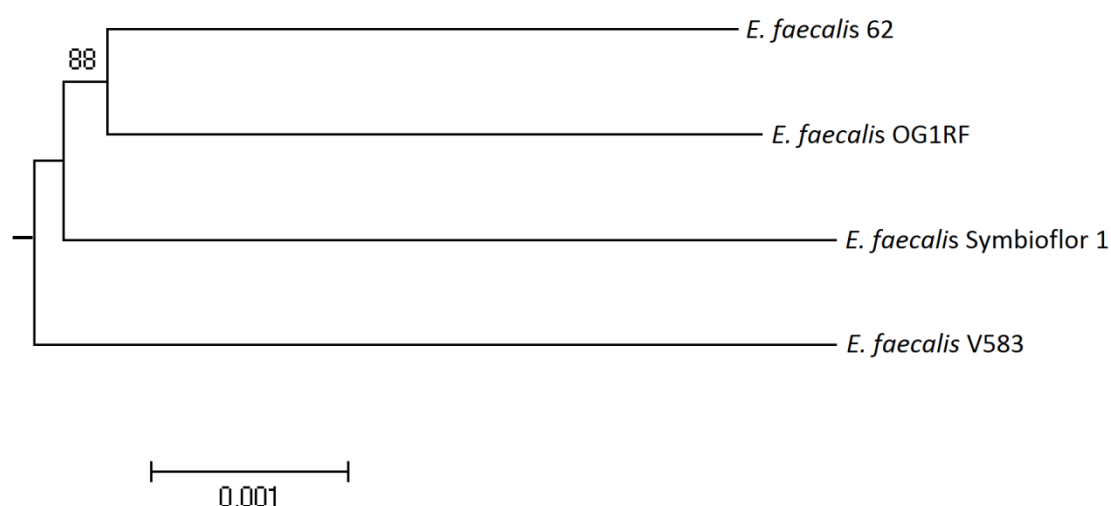


Abbildung 15: Phylogenetischer Baum der 4 *E. faecalis*-Stämme.

Erstellt mit der neighbour-joining-Methode und einem Bootstrap-Value von 100, MEGA5

Es zeigt sich, dass die 4 Stämme genetisch relativ nahe verwandt sind und jeweils etwa gleich weit voneinander entfernt. Anhand der Berechnungen entstammen *E.*

faecalis 62 und OG1RF einem gemeinsamen Vorgänger, welcher wiederum mit Symbioflor 1 einen gemeinsamen Vorläufer hat, und dieser hat einen gemeinsamen Vorläufer mit V583.

Dies ist übereinstimmend mit dem phylogenetischen Baum, der auf der Basis des MLSTs erstellt wurde. Der Baum, der auf Basis des Kerngenoms erstellt wurde, gleicht die Abstände der Stämme zueinander etwas an. Dies ist nicht verwunderlich, wenn man bedenkt, dass durch die große Datenmenge einzelne Unterschiede vermutlich weniger ins Gewicht fallen.

4.5 Sequenz-Vergleich der Genome

Einen ersten vergleichenden Überblick über die Genome bietet eine Ausrichtung aller 4 Genome zueinander mit MAUVE:

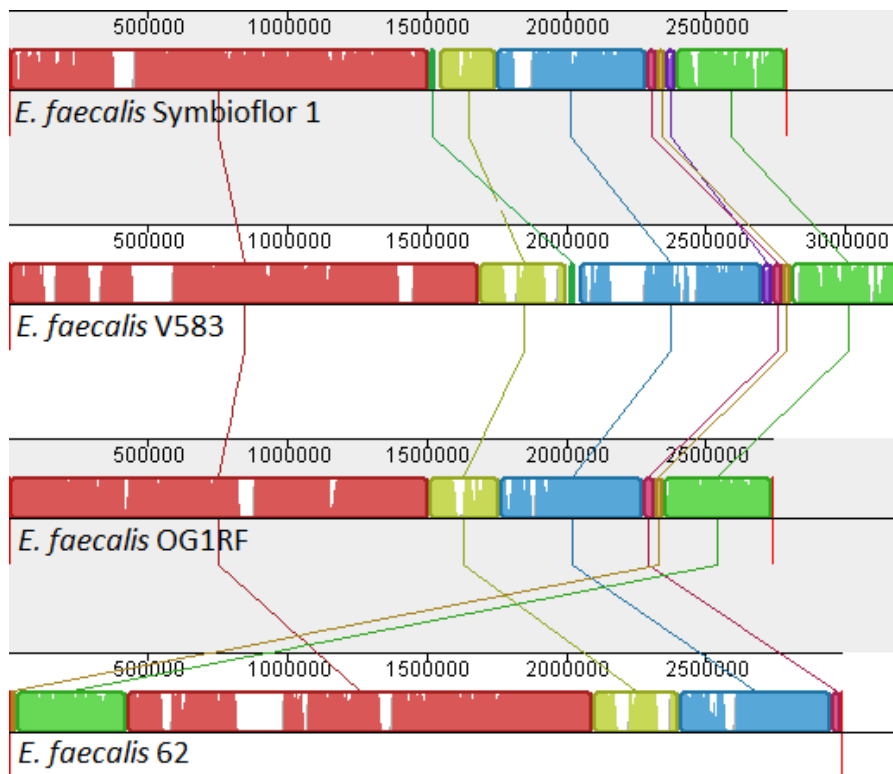


Abbildung 16: Homologie-Vergleich der *E. faecalis*-Stämme mit MAUVE.

Homologe lokale colineare Blöcke in der gleichen Farbe. Von oben nach unten: *E. faecalis* Symbioflor 1, *E. faecalis* V583, *E. faecalis* OG1RF und *E. faecalis* 62

Die Stämme sind sich vordergründig sehr ähnlich. Auffallend ist, dass über weite Stellen der Genome sehr hohe Sequenzhomologien vorhanden sind, die an manchen Stellen eher abrupt durch große spezifische Abschnitte unterbrochen sind.

Das einzige größere Rearrangement betrifft die Region Ef1_2117 bis Ef1_2184 in *E.faecalis* Symbioflor 1, die sonst nur in *E. faecalis* V583 in den Genen EF2856 bis EF2925 ihre Entsprechung findet. Diese etwa 34 kbp lange Sequenz gehört zu einer von PHAST als Prophagen indentifizierten Region (Zhou et al. 2011).

4.5.1 Synteny der Codierenden Sequenzen (CDS)

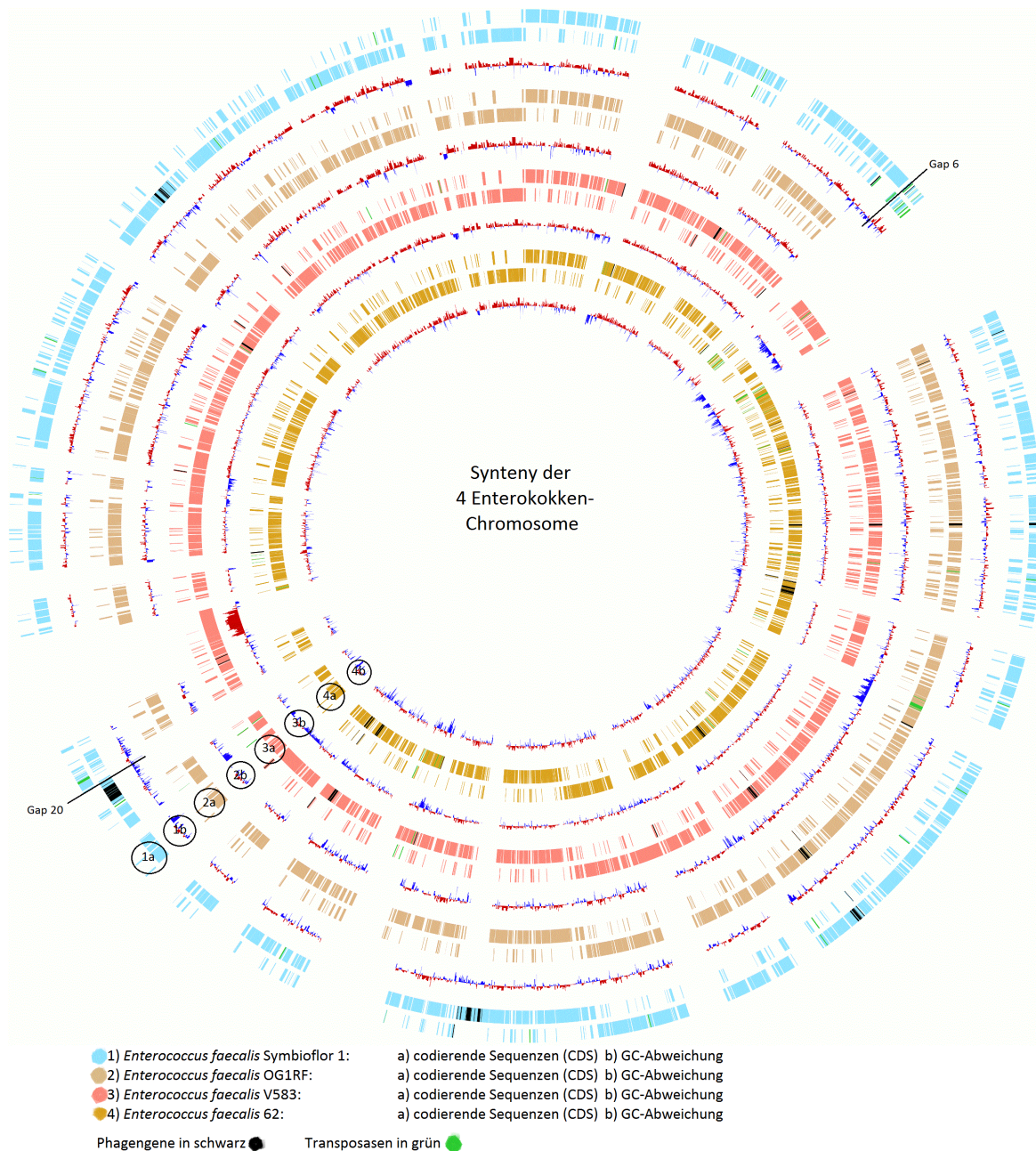


Abbildung 17: Pangenom der proteinkodierenden Sequenzen der 4 Enterokokken-Stämme

Das Bild wurde erstellt in GenomeViz (Ghai et al. 2004). Die syntene Ausrichtung der Genome erfolgte mit GeCo (C. T. Kuenne et al. 2007).

4.5.2 Statistische Analyse der genetischen Homologie

Folgende Tabelle stellt dar, wie viele Gene für jeden Organismus spezifisch sind, wie viele von mehreren und wie viele von allen geteilt werden. Für jede Kombination zwischen den Organismen ist angegeben, wie viele Cluster existieren und wie viele

Gene aus welchem Organismus sich in den jeweiligen Clustern befinden. Ein einzelnes Cluster besteht hierbei aus einem Gen und all seinen Homologen in allen betreffenden Organismen, das heißt, es zählt auch Kopien eines Genes in ein und demselben Organismus. Daher ist die Zahl der Gene, die ein Organismus in den jeweiligen Clustern hat, oft höher als die Anzahl der Cluster der entsprechenden Kombination (C. T. Kuenne et al. 2007).

Homology Matrix aus GeCo mit 80% Identität, 90% Abdeckung (Aminosäuren)					
	Cluster	Gene			
		<i>E. faecalis</i> Symbioflor 1	<i>E. faecalis</i> V583	<i>E. faecalis</i> OG1RF	<i>E. faecalis</i> 62
alle	2008	2013	2019	2015	2017
alle außer Symbioflor 1	61		62	61	62
alle außer 62	63	63	63	63	
alle außer V583	101	103		101	102
alle außer OG1RF	69	70	69		75
Symbioflor 1 und 62	79	83			94
Symbioflor 1 und V583	43	48	51		
Symbioflor 1 und OG1RF	12	12		12	
62 und V583	105		113		109
62 und OG1RF	50			50	50
V583 und OG1RF	44		44	44	
nur Symbioflor 1	306	324			
nur 62	378				388
nur V583	673		691		
nur OG1RF	233			233	

Tabelle 4: Homologie-Matrix der 4 *E. faecalis*-Genome. Output aus GeCo

Legt man also 80% Identität und 90% Abdeckung der Aminosäuresequenz als Grenze fest, besteht das Kerngenom der 4 *E. faecalis*-Stämme aus 2008 proteinkodierenden Sequenzen. Jeder Stamm hat für ihn spezifische Gene: Die meisten hat *E. faecalis* V583 mit 691, gefolgt von *E. faecalis* 62 mit 388, *E. faecalis* Symbioflor 1 mit 324 und *E. faecalis* OG1RF mit 233.

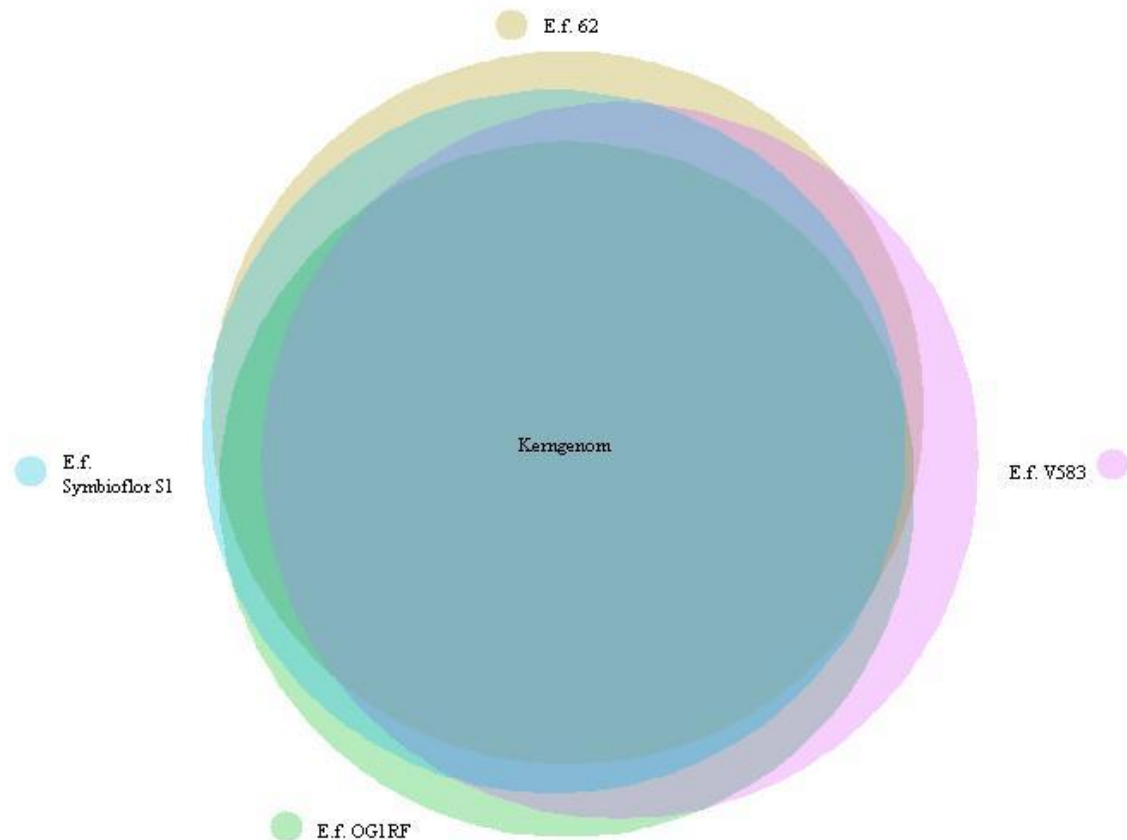


Abbildung 18: Euler-Diagramm zur Visualisierung der Homologie

Als Grundlage dient die Homologie-Matrix aus GeCo mit 80% Identität und 90% Abdeckung. Die Flächen sind annähernd proportional der Anzahl an Genen in den entsprechenden Gruppen. Erstellt mit venneuler in R (Wilkinson 2012).

Bei niedrigeren Grenzwerten von 60% Identität und 80% Abdeckung steigt die Anzahl der Kerngene auf 2081, die spezifischen Gene sind dementsprechend weniger: 601 in *E. faecalis* V583, 330 in *E. faecalis* 62, 257 in *E. faecalis* Symbioflor 1, und 179 in *E. faecalis* OG1RF.

Bei noch niedrigeren Grenzwerten von 50% Identität und 40% Abdeckung bilden 2105 Gene das Kerngenom, *E. faecalis* V583 hat 553 spezifische Gene, *E. faecalis* 62 hat 257, *E. faecalis* Symbioflor 1 hat 191 und *E. faecalis* OG1RF hat 135.

Die Zahl der Gene im Kerngenom ändert sich demnach nur in geringem Maße mit Änderung der zugrunde liegenden Grenzwerte. Dies deckt sich mit der Beobachtung im globalen Sequenzvergleich der Genome, die über weite Strecken hohe Homologien zeigte, die von kurzen spezifischen Regionen unterbrochen sind. Demnach ist in der Mehrzahl ein Gen entweder homolog in den verschiedenen Organismen oder spezifisch für einen Organismus, Homologien mit niedrigem Übereinstimmungsgrad kommen eher selten vor.

4.5.3 COG-Klassifikation der Gene

Die CDS wurden in Augur nach COG-Gruppen eingeteilt (Tatusov 2000; André Billion et al. 2006). Eine Übersicht über die Anzahl der Gene in den verschiedenen COG-Gruppen gibt folgende Tabelle:

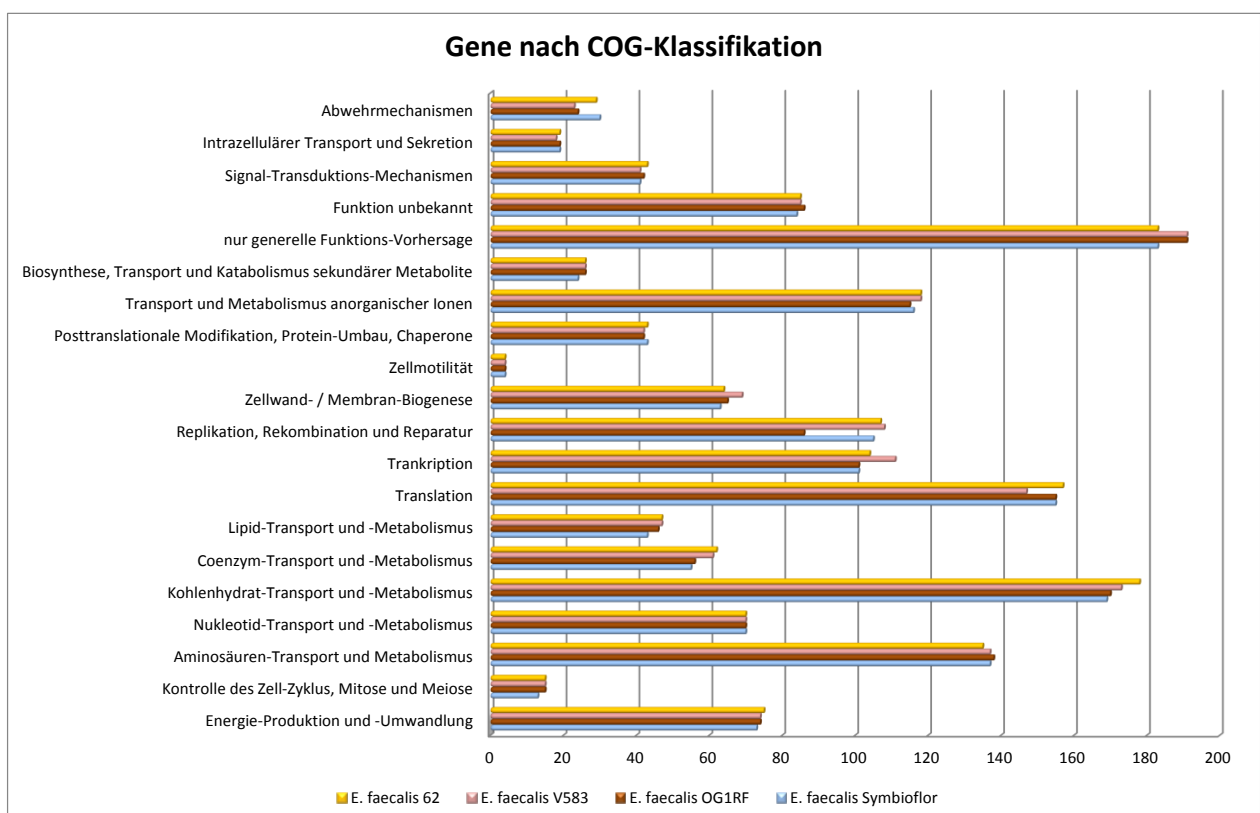


Abbildung 19: Klassifikation der Gene nach der COG-Klassifikation

In der Einteilung nach der COG-Klassifikation sind sich die Organismen recht ähnlich. *E. faecalis* V583 und *E. faecalis* OG1RF verfügen über weniger Gene für Abwehrmechanismen. *E. faecalis* OG1RF verfügt über deutlich weniger Gene des Bereichs Replikation, Rekombination und Reparatur. *E. faecalis* V583 besitzt mehr

Gene, die die Zellwand-Biosynthese betreffen, als die anderen Stämme. *E. faecalis* Symbioflor 1 und *E. faecalis* OG1RF besitzen im Vergleich weniger Gene für Kohlenhydrat-Transport und –Metabolismus, ebenso für Coenzym-Transport und –Metabolismus.

4.5.4 Spezifische Gene des *Enterococcus faecalis* Symbioflor 1

Mittels GECO wurde ermittelt, welche Gene nur *E. faecalis* Symbioflor 1, und keiner der 3 anderen Stämme enthält:

- Von den 189 für *E. faecalis* Symbioflor 1 spezifischen Genen sind 19 vermutlich Pseudogene. Für sie existiert eine längere Version in den Referenzstämmen.
- 93 spezifische Gene liegen im Bereich von Phagen oder Transposons.
- 31 Gene sind als hypothetical protein annotiert, 60 als conserved hypothetical protein und 2 als conserved domain protein.
- 6 Gene sind Bestandteile von PTS-Systemen.
- 5 Gene sind Transkriptionsregulatoren, 1 davon sehr wahrscheinlich.
- 4 Gene codieren vermutlich sezernierte Proteine.
- 2 Gene gehören vermutlich zu DNA-Methyltransferasen.
- 2 Gene gehören zu einem ABC-Transporter, eines davon zu einem Multidrug-ABC-Transporter-System.

4.5.4.1 Lactose/Cellobiose-spezifisches PTS-System

Die Region von EFS1_0028 bis EFS1_0031 in *E. faecalis* Symbioflor 1 beherbergt 4 Gene, die in den anderen 3 Stämmen nicht enthalten sind. Eine entsprechende Region mit zweien dieser Gene und hoher Homologie auf Nukleotid-Ebene findet sich bei *Enterococcus sp.* 7L76.

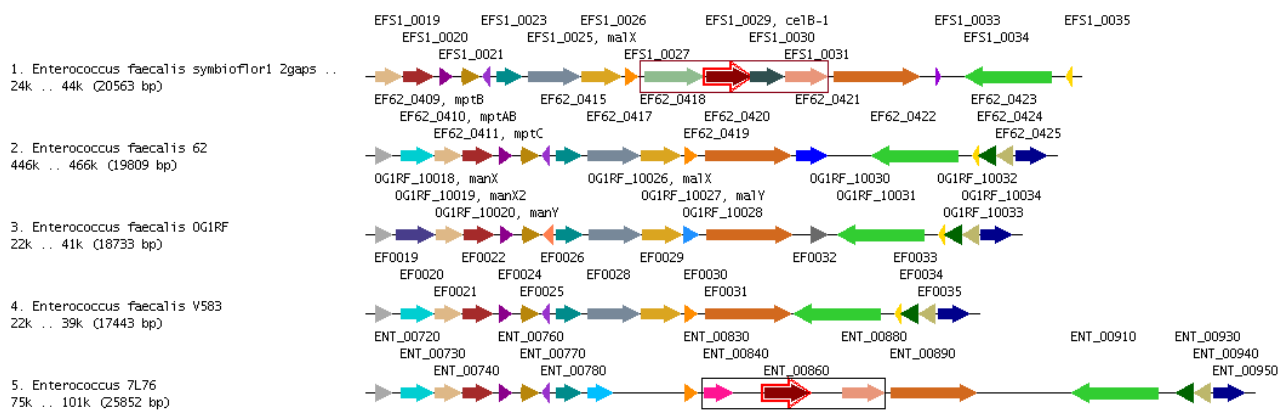


Abbildung 20: Lactose/Cellobiose-spezifisches PTS-System

Homologe CDS jeweils in der gleichen Farbe

Während der Analyse wurden öfters Übereinstimmungen von *E. faecalis* Symbioflor 1 zu *Enterococcus sp.* 7L76 gefunden. Dessen Genom ist verfügbar auf der Webseite des NCBI und wurde als Referenzgenom für das metaHIT Projekt sequenziert (Pajon et al. 2010) .

Die offenen Leseraster der Gene unterscheiden sich allerdings im Vergleich: Das erste Gen ist in *E. faecalis* Symbioflor 1 (conserved hypothetical protein) deutlich länger als in *Enterococcus sp.* 7L76 (transcriptional antiterminator). Das zweite Gen ist homolog (PTS system, lactose/cellobiose-specific IIC component). Das dritte Gen ist nur in *E. faecalis* Symbioflor 1 enthalten (sugar isomerase (SIS) domain protein). Das vierte wiederum ist homolog in beiden Stämmen (glycosyl hydrolase, family 4 protein).

4.5.4.2 Potentielles Insertionselement mit Restriktionssystem und Zucker-Metabolismus-Genen

Die Region von EFS1_2420 bis EFS1_2431 im Genom von *E. faecalis* Symbioflor 1 beschreibt einen Locus, welcher in den drei Referenz-Enterokokken nicht enthalten ist. Eine annähernd gleiche Sequenz findet sich jedoch wiederum in *Enterococcus* sp. 7L76.

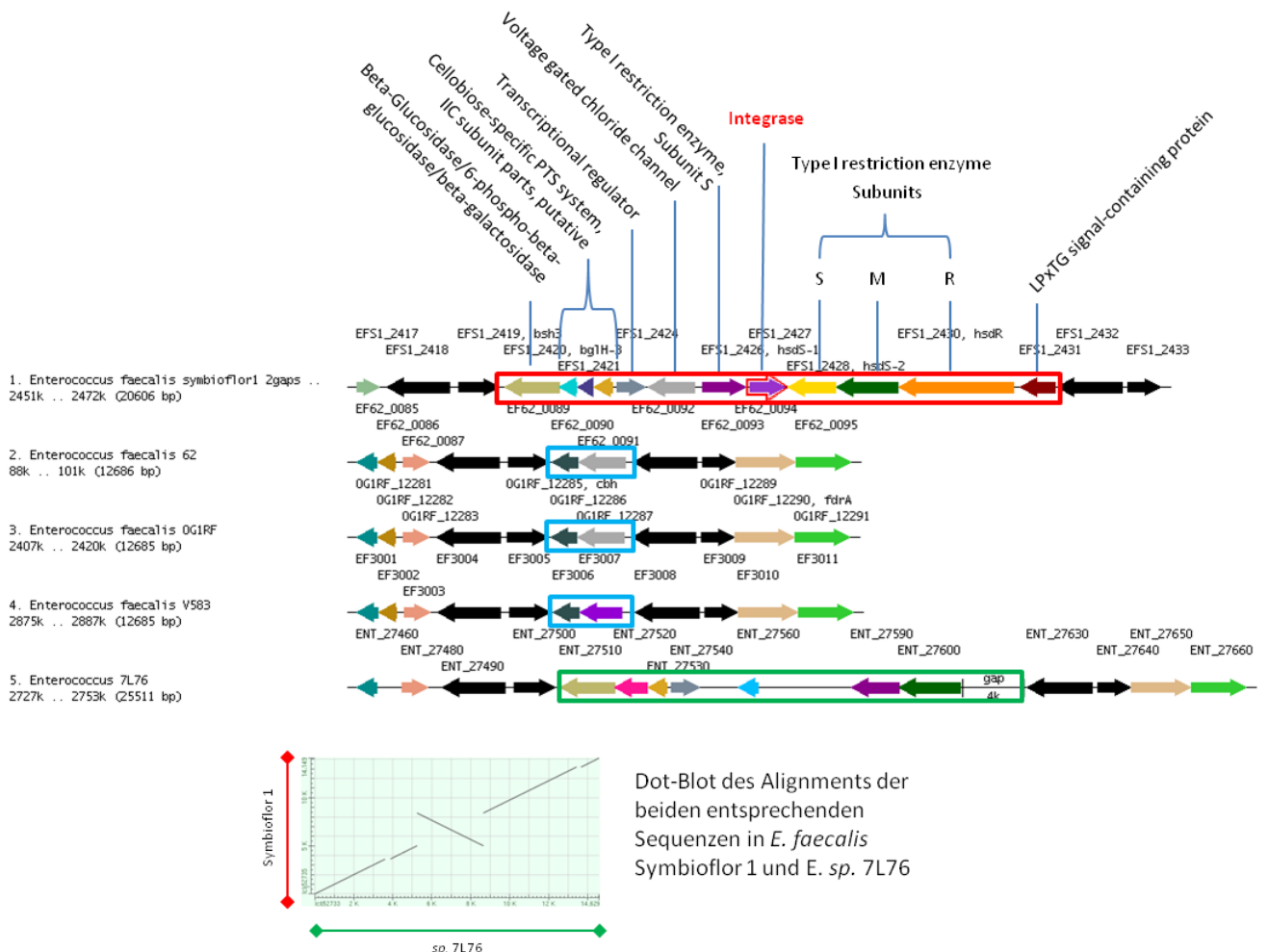


Abbildung 21: Potentielles Insertionselement mit Restriktionsenzym und Genen für Zucker-Metabolismus

Oben: GeCo (80% Identität, 90% Abdeckung). Homologe CDS jeweils in der gleichen Farbe, konservierte CDS in schwarz
Unten: Dot-Alignment der beiden Sequenzen, erstellt auf der Homepage des NCBI

Der Dot-Plot zeigt, anders als der Vergleich auf Gen-Ebene, eine sehr hohe Homologie der beiden Sequenzen in *E. faecalis* Symbioflor 1 und *Enterococcus sp.* 7L76, lediglich in der Mitte ist eine Inversion zu finden. Die Gene im linken Bereich codieren für ein Cellobiose-spezifisches PTS-System (IIC-Teil), jedoch ist dieses in *E. faecalis* Symbioflor 1 geteilt. Die codierende Sequenz ganz rechts in *E. faecalis* Symbioflor 1 beinhaltet ein LPxTG-Motiv, ein Hinweis auf ein Oberflächenprotein (Hendrickx et al. 2009).

Die Stämme *E. faecalis* V583, *E. faecalis* OG1RF und *E. faecalis* 62 besitzen von dem bezeichneten Locus nur einen Chloridkanal und ein hypothetisches Protein.

4.5.4.3 Multidrug-transporter

Eine Multidrug-ABC-transporter-CDS findet sich wiederum nur in *E. faecalis* Symbioflor 1 und in *Enterococcus sp.* 7L76.

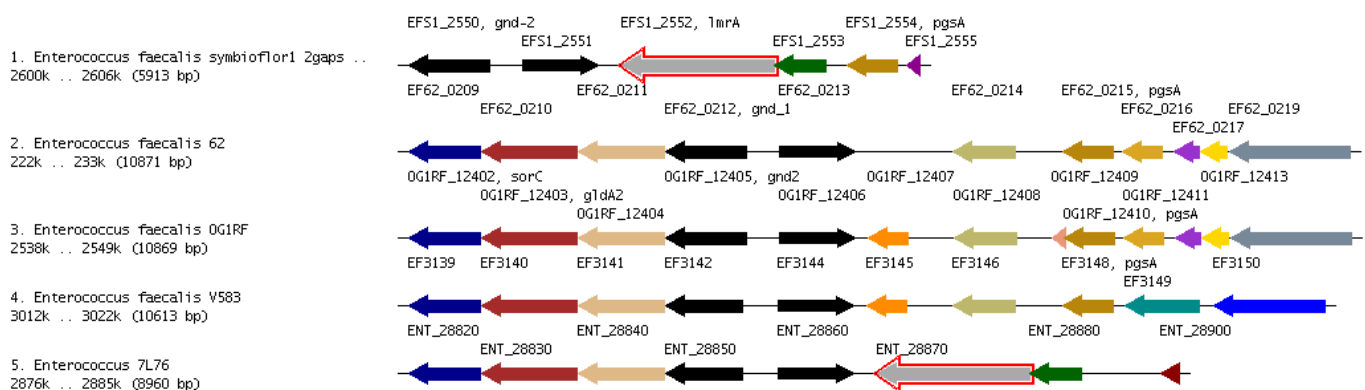


Abbildung 22: Region um den Multidrug-ABC-transporter

in GeCo (80% Identität, 90% Abdeckung)

4.5.4.4 Spezifische DNA-Sequenzen des *E. faecalis* Symbioflor 1

Mittels des Programms ssgenefinder (Ho et al. 2011) wurden Sequenzen im Genom von *E. faecalis* Symbioflor 1 gesucht, die für diesen spezifisch sind. Um unspezifische Sequenzen zu identifizieren, wurden dem Programm die Genome von *E. faecalis* V583, *E. faecalis* 62 und *E. faecalis* OG1RF als Referenzen eingegeben.

Die als unique identifizierten Sequenzen wurden gegen die nucleotide collection des NCBI geblastet. (<http://blast.ncbi.nlm.nih.gov>). Hierbei zeigte sich, dass viele der Sequenzen starke Homologien zu *Enterococcus sp.* 7L76 hatten. Dessen Draft-

Genom wurde deshalb in einem neuen Lauf der ssgenefinder-Software als zusätzliche Referenz zur Verfügung gestellt.

Die resultierenden 92 Sequenzen von insgesamt 105.177 bp wurden erneut gegen die nucleotide collection geblastet (Option megablast): Übrig blieben 28 ausschließlich in *E. faecalis* Symbioflor 1 vorhandene Sequenzen mit einer Durchschnittslänge von 747 bp, insgesamt 20.904 bp.

4.5.5 Putative Pathogenitätsfaktoren

4.5.5.1 Aggregation Substance

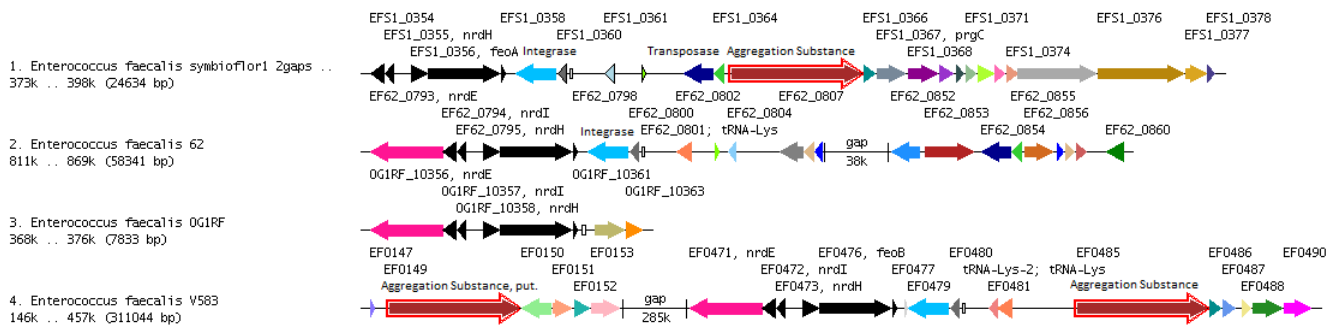


Abbildung 23: Gene im Bereich der Aggregation Substance

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Das entsprechende Gen findet sich einfach in *E. faecalis* Symbioflor 1 (EF1_0365) und doppelt in *E. faecalis* V583 (EF0149 und EF0485). In den anderen beiden Stämmen ist es nicht enthalten

Die Region um Aggregation Substance unterscheidet sich zwischen beiden Stämmen. Bei beiden ist eine Integrase in der Nähe zu finden, bei *E. faecalis* Symbioflor 1 zudem eine Transposase.

4.5.5.2 Collagen-Adhesion-Protein

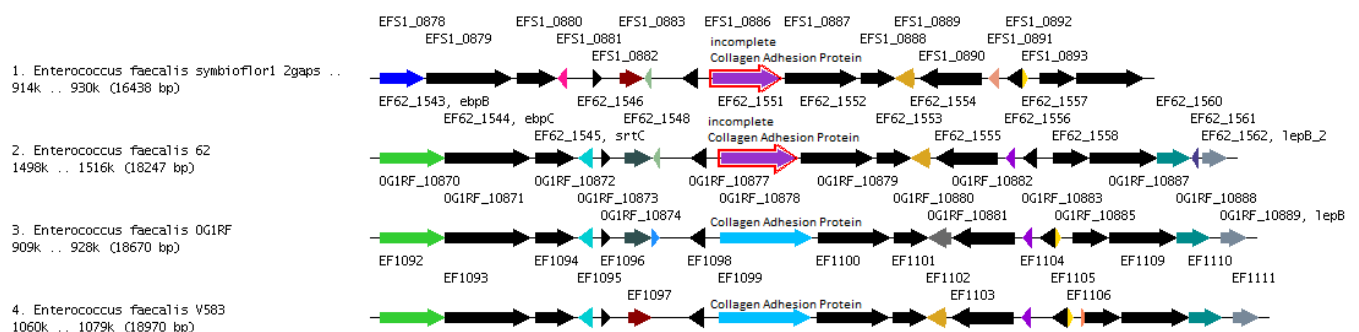


Abbildung 24: Gene im Bereich des Collagen Adhesion Protein

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Dieses Protein ist in allen 4 Stämmen enthalten. Die Region ist relativ stark konserviert.

Allerdings ist das Gen in *E. faecalis* Symbioflor 1 und *E. faecalis* 62 im Vergleich zu den anderen beiden Stämmen verkürzt.

4.5.5.3 Gelatinase

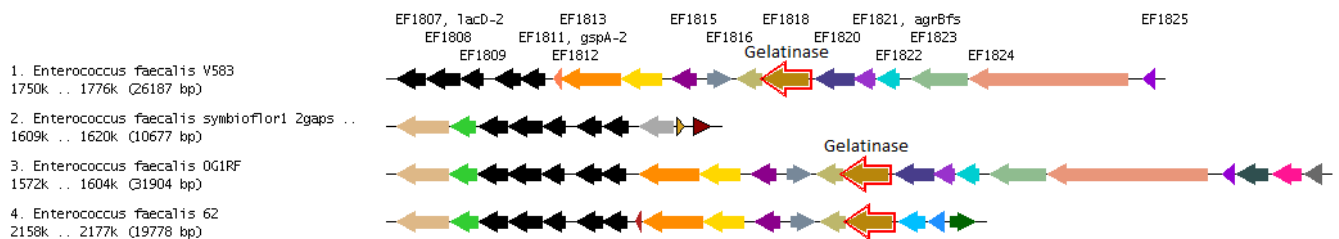


Abbildung 25: Gene im Bereich der Gelatinase

Erstellt in GeCO (80% Identität, 90% Abdeckung)

Das Protein Gelatinase (auch Coccolysin) ist in allen Stämmen außer *E. faecalis* Symbioflor 1 in einer relativ homologen Region vorhanden (EF1818, EF62_2189, OG1RF_11526). Es handelt sich um eine etwa 24 kbp lange Region, die in *E. faecalis* Symbioflor 1 fehlt. An deren Stelle sind Transposasen-Gene zu finden.

4.5.5.4 Cytolysin L

Das funktionelle Cytolysin L besteht aus einer langen und einer kurzen Untereinheit (Tendolkar et al. 2004). Beide sind in *E. faecalis* V583 vorhanden (EF0525, EF0526), im Gegensatz zu den anderen drei untersuchten Stämmen.

4.5.5.5 Enterococcus-Surface-Protein

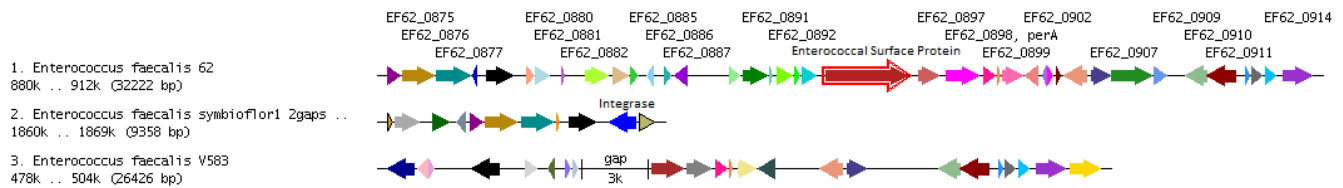


Abbildung 26: Gene im Bereich des Enterococcal Surface Proteins

Erstellt in GeCO (80% Identität, 90% Abdeckung)

Das Gen für das Enterococcus-Surface-Protein *esp/efaA* ist nur im Genom von *E. faecalis* 62 (EF62_0896) enthalten. Bei *E. faecalis* Symbiflor 1 ist in der Umgebung des Locus eine potentielle Integrase zu finden.

4.5.5.6 Pathogenitätsinsel 1

Die Gene für Aggregation Substance und Cytolysin-L liegen bei *E. faecalis* V583 in einer gemeinsamen Region mit im Vergleich zum Restgenom niedrigem GC-Gehalt (32%), die als Pathogenitätsinsel bezeichnet wurde (Laverde Gomez, et al., 2011).

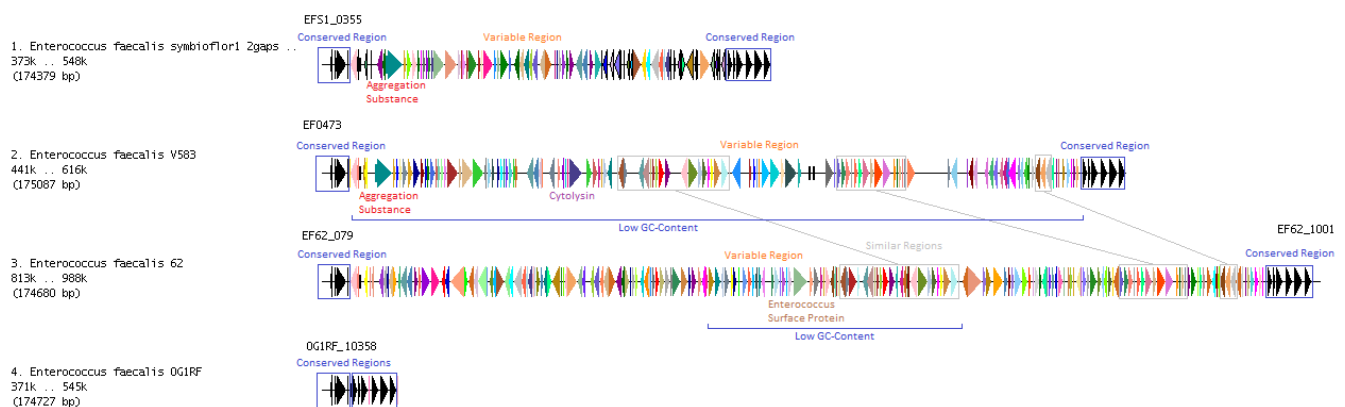


Abbildung 27: Pathogenitätsinsel 1

Enthält Aggregation Substance, Cytolysin und Enterococcus Surface Protein. Erstellt in GeCo (80% Identität, 90% Abdeckung).

Sowohl *E. faecalis* Symbioflor 1 als auch *E. faecalis* 62 besitzen an der entsprechenden Stelle in ihrem Genom eine ähnliche Region. *E. faecalis* Symbioflor 1 besitzt im stromaufwärts liegenden Teil dieser Sequenz das Gen für Aggregation Substance, während *E. faecalis* 62 stromabwärts Homologien zu *E. faecalis* V583 aufweist, nahe daran liegt auch das Gen für das Enterococcus-Surface-Protein. Alle drei Stämme besitzen in dieser Region Transposasen-Gene und einen niedrigen GC-Gehalt. In *E. faecalis* OG1RF ist die entsprechende Region nicht vorhanden.

4.5.5.7 Hyaluronidase

In keinem der Stämme konnte ein Gen für Hyaluronidase gefunden werden.

4.5.5.8 Pathogenitätsinsel 2: Vancomycin-Resistenz, Bacteriocin

Nur *E. faecalis* V583 hat in seinem Genom das Gen für Bacteriocin aufzuweisen (EF2314). Es liegt in einer Region mit überdurchschnittlich hohem GC-Gehalt (43%), welche ebenfalls als Pathogenitätsinsel bezeichnet wurde (Domann et al. 2007; Vebø et al. 2010). Sie enthält Rekombinase- und Integrase-Gene.

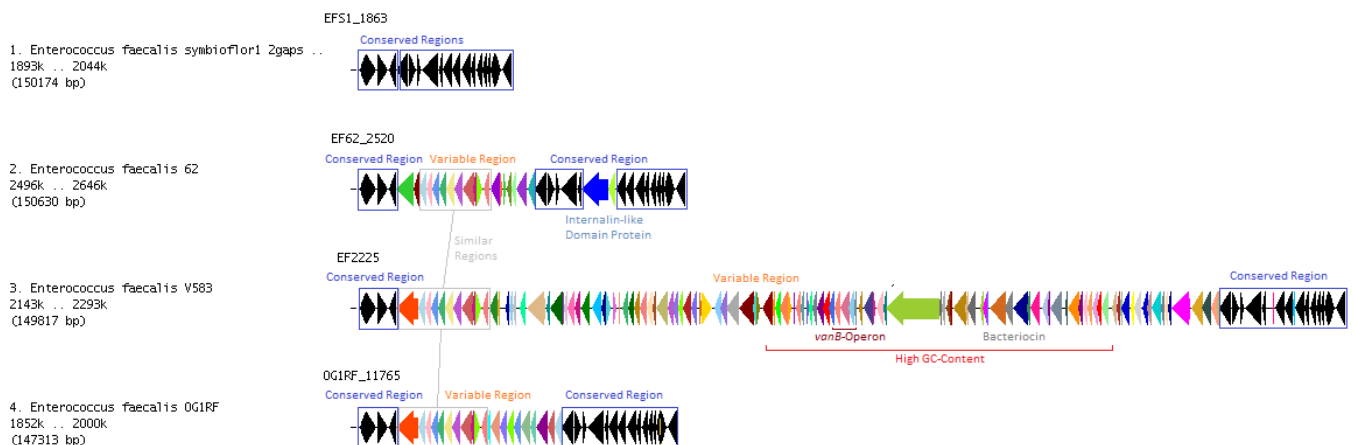


Abbildung 28: Pathogenitätsinsel 2

Enthält *vanB*-Operon und Bacteriocin. Erstellt in GeCo (80% Identität, 90% Abdeckung)

Die GC-reiche Region ist nur in *E. faecalis* V583 vorhanden. Der Locus im gesamten scheint durch hohe Varianz ausgezeichnet zu sein, einzig eine Übereinstimmung weniger CDS lässt sich zwischen den Stämmen *E. faecalis* 62, *E. faecalis* OG1RF

und *E. faecalis* V583 finden. In *E. faecalis* Symbioflor 1 fehlt der variable Teil dieses Locus vollständig.

In dieser Region beherbergt *E. faecalis* V583 das *vanB*-Operon, welches die Resistenz gegenüber Vancomycin verleiht. Dieses fehlt somit in *E. faecalis* Symbioflor 1 und den anderen beiden Stämmen (Patrice Courvalin 2006; Dahl et al. 1999).

4.5.5.9 *vanG*-ähnlicher Locus

Es gibt einen Locus, der dem *vanG*-Operon ähnelt. Er ist in allen vier untersuchten Stämmen bis auf *E. faecalis* V583 enthalten.

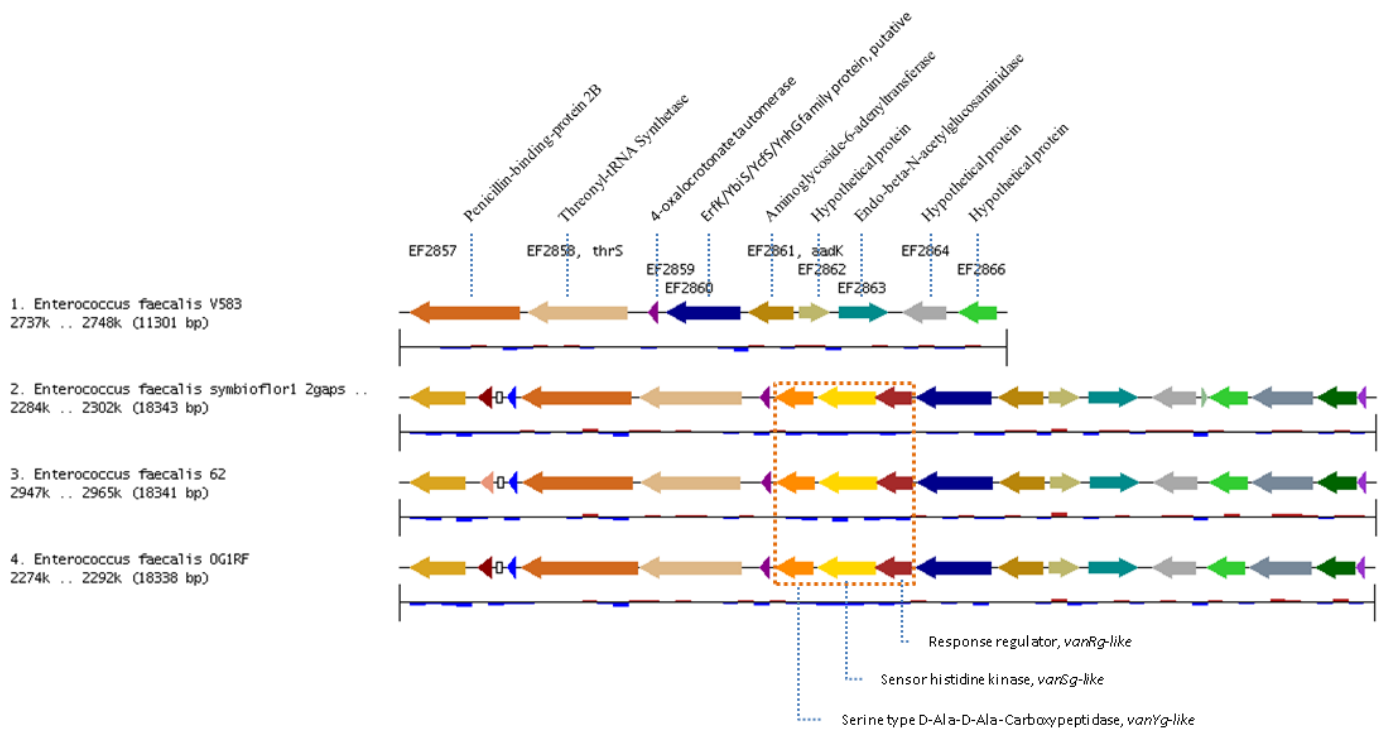


Abbildung 29: *vanG*-ähnlicher Locus

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Der Locus weist zwar Homologien zum Vancomycin-Resistenz-Operon *vanG* auf, jedoch fehlen zur Funktionalität entscheidende Proteine. Vermutlich kodiert dieser Locus einen bisher unbekannten regulatorischen Stoffwechselweg (Bourgogne et al. 2008).

4.5.5.10 Gene zur Verteidigung gegen oxidativen Stress

Die Gene, die den oxidativen Stress betreffen, unterliegen durchweg einem hohen Grad an Homologie zwischen den 4 Bakterienstämmen. Die folgende Tabelle listet die Gene in den 4 Genomen auf.

Genprodukt	<i>E. faecalis</i> Symbioflor 1	<i>E. faecalis</i> V583	<i>E. faecalis</i> 62	<i>E. faecalis</i> OG1RF
Superoxid-Dismutase	EF1_0345	EF0463	EF62_0785	OG1RF_10348
Katalase	EF1_1311/EF1_1312	EF1597	EF62_1974	OG1RF_11314
Glutathion-Reduktase	EF1_2657	EF3270	Ef62_0323	OG1RF_12519
NADH-Peroxidase	EF1_0994	EF1211	EF62_1655	OG1RF_10983
NADH-Oxidase (1)	EF1_1302	EF1586	EF62_1966	OG1RF_11305
Thioredoxin-Reduktase (1)	EF1_1118/EF1_1119	EF1338	EF62_1791	OG1RF_11128
Thioredoxin-Reduktase (2)	EF1_2181	EF2738	EF62_2892	OG1RF_12102
Peroxiredoxin	EF1_2182	EF2739	EF62_2893	OG1RF_12103
NADH-Oxidase (2)	EF1_1621	EF1932	EF62_2298	OG1RF_0311
NADH-Dehydrogenase (1)	EF1_2647	EF3257	EF62_0311	OG1RF_12510
NADH-Dehydrogenase (2)	EF1_1684	EF2055	EF62_2417	OG1RF_11660

Tabelle 5: Gene zur Verteidigung gegen oxidativen Stress

Auffallend ist, dass die Gene der Katalase sowie der Thioredoxin-Reduktase (1) im Genom von *E. faecalis* Symbioflor 1 geteilt sind.

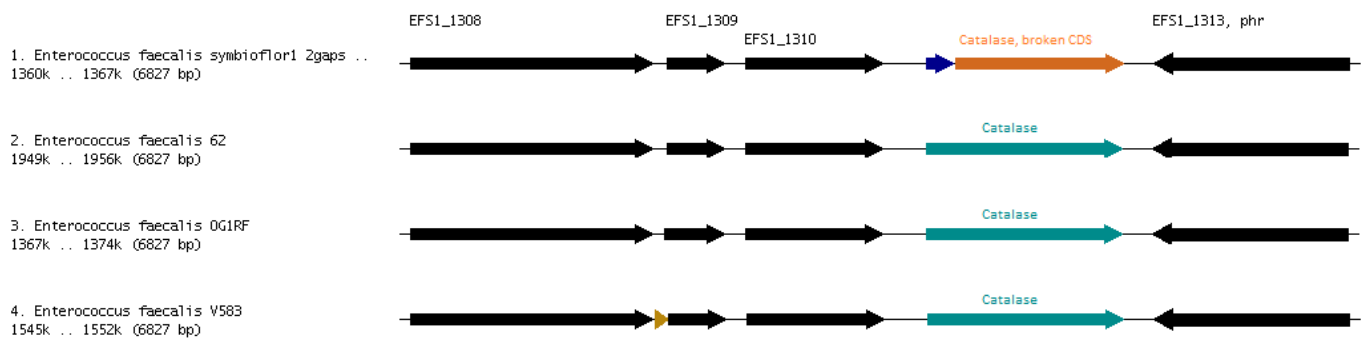


Abbildung 30: Katalase - fragmentiertes Gen

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Im Gen der Katalase ist bei *E. faecalis* Symbioflor 1 durch einen SNP ein vorzeitiges Stop-Codon vorhanden. Kurz darauf folgt ein Startcodon, sodass vermutlich ein im Vergleich zu den Referenzstämmen fragmentiertes Protein entsteht.

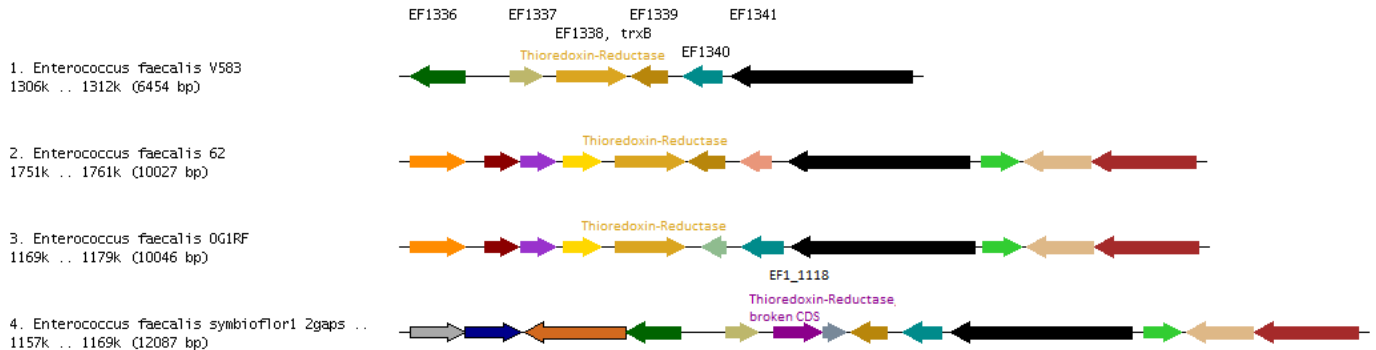


Abbildung 31: Thioredoxin-Reduktase - fragmentiertes Gen

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Bei der Thioredoxin-Reduktase (1) des *E. faecalis* Symbioflor 1 ist durch eine Deletion eine Verschiebung des Leserasters entstanden, wodurch ein vorzeitiges Stopcodon vermutlich ebenfalls zur Verkürzung des Proteins führt.

Abgesehen von den gezeigten sind die Regionen mit Genen für die Abwehr von oxidativem Stress auffallend homolog.

4.5.5.11 Kapsel

Zwei Loci sind für die Biosynthese einer Polysaccharid-Kapsel in Betracht gezogen worden (Thurlow et al. 2009).

1. Das *cps*-Operon

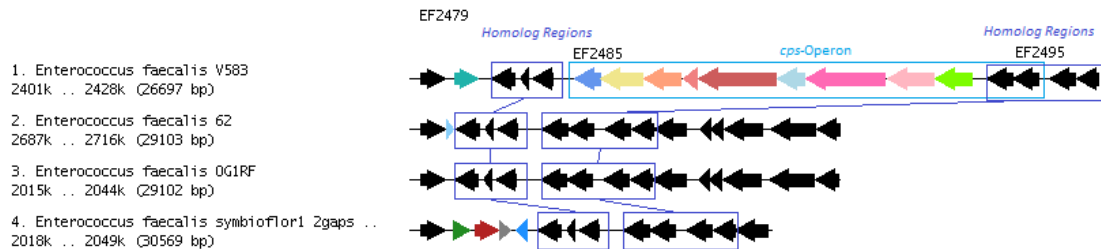


Abbildung 32: Das *cps*-Operon

Erstellt in GeCo (80% Identität, 90% Abdeckung)

Von den gezeigten besitzt einzig *E. faecalis* V583 alle 11 ORFs des *cps*-Operons. Die anderen drei Stämme besitzen nur die ersten 2 ORFs (Hancock & Gilmore 2002).

2. Das *epa*-Operon

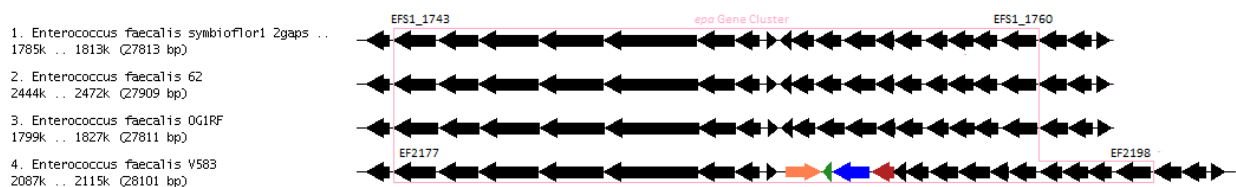


Abbildung 33: Das *epa*-Operon

Erstellt in Geco (80% Identität, 90% Abdeckung)

Im Gegensatz dazu ist das *epa*-Operon in allen untersuchten Stämmen mit seinen gesamten 16 ORFs komplett enthalten. *E. faecalis* V583 enthält in diesem Locus zusätzlich 4 weitere ORFs (Xu et al. 1998; Teng et al. 2009).

4.5.5.12 CRISPR Loci

CRISPR-Systeme stellen eine Verteidigung prokaryontischer Genome gegen fremde DNA aus Phagen oder Plasmiden dar. Zu einem funktionsfähigen System gehört eine Reihe von kurzen palindromischen repeats, die von ebenfalls kurzen spezifischen spacern unterbrochen sind. Die spezifischen spacer bestimmen, gegen welche fremde DNA sich das CRISPR-System richten kann. Zur Funktionalität sind zusätzlich Endonukleasen erforderlich, die von den assoziierten *cas*-Genen codiert werden (Palmer & Gilmore 2010; Bourgoigne et al. 2008; Barrangou et al. 2007).

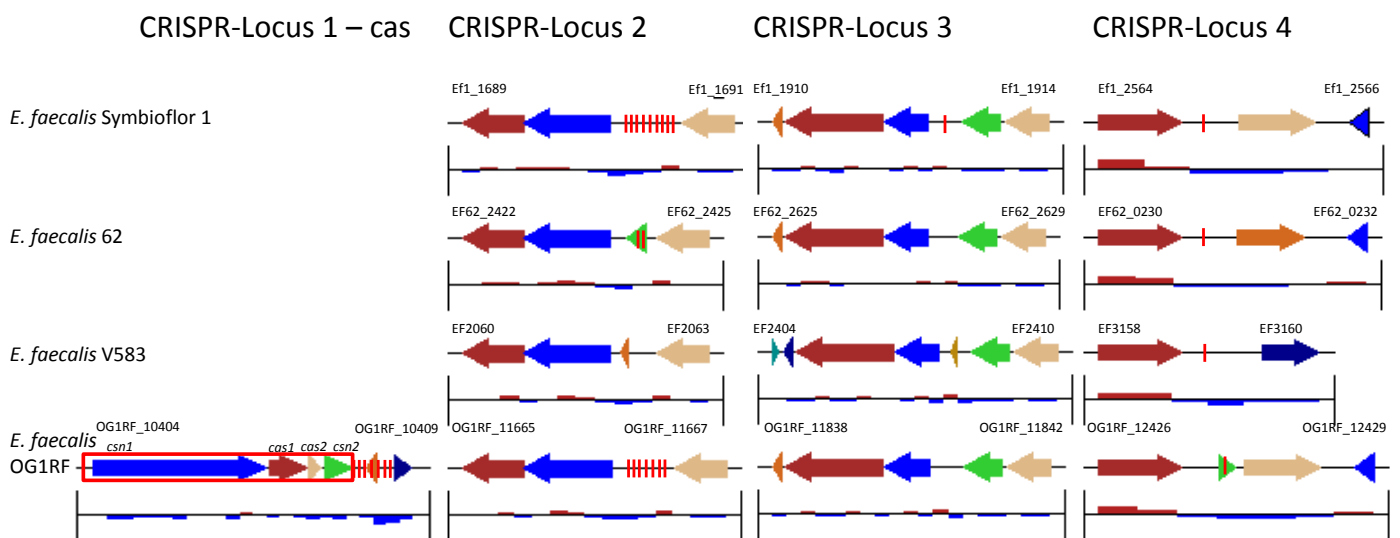


Abbildung 34: CRISPR-Loci

Mit umgebenden CDS, modifiziert aus GeCo (80% Identität, 90% Abdeckung)

Im Bereich des jeweiligen Locus homologe Gene sind in der gleichen Farbe dargestellt. Die repeats sind als rote senkrechte Striche eingezeichnet. Unter der Region ist jeweils die Abweichung des GC-Gehalts von der Norm angegeben. Zur Orientierung ist der Locustag jeweils des ersten und letzten gezeigten CDS jedes Locus angegeben.

Die CRISPR Loci wurden mit CRISPRFinder detektiert (Grissa et al. 2007).

Von den untersuchten Stämmen verfügt nur *E. faecalis* OG1RF über einen CRISPR-Locus mit assoziierten *cas*-Genen.

Ein weiterer Locus ohne assoziierte *cas*-Gene ist in *E. faecalis* OG1RF, *E. faecalis* Symbioflor 1 und *E. faecalis* 62 vorhanden, unterscheidet sich allerdings in Anzahl der repeats und Spezifität der spacer.

E. faecalis Symbioflor 1 verfügt über zwei weitere mögliche Loci, die jedoch nur über 1 repeat und keine *cas*-Gene verfügen. Einer dieser Loci ist in allen vier Stämmen homolog vorhanden.

Eine Überprüfung der spacer mittels Blast in der nucleotide collection des NCBI ergab bis auf zwei Ausnahmen keine Ergebnisse: Der vierte spacer des ersten CRISPR-Locus von *E. faecalis* Symbioflor 1 beinhaltet Sequenz von *Enterococcus* phage phiFL3B, während der dritte spacer des zweiten CRISPR-Locus von *E. faecalis* OG1RF zu einem *Enterococcus faecalis*-Plasmid passt (pLG2-0010).

4.6 Die verbleibenden Lücken

Im aktuellen Assembly gibt es zwei Contigs mit den Größen 1.434.252 bp sowie 1.355.769 bp, die am Genom von *E. faecalis* V583 gemappt sind. Dazwischen liegen Gap 6 und 20. Zusätzlich gibt es ein Contig mit der Größe 8.016 bp, welches ungemappt ist, also in einem der beiden Gaps liegen könnte.

4.6.1 Übersicht über Gap 6

4.6.1.1 PCRs

Für Gap 6 konnten PCRs generiert werden, die nach elektrophoretischer Kontrolle als ausreichend gut für eine Sequenzierung betrachtet wurden. Anfänglich waren auch die Reads gut. Das Primerwalking scheiterte jedoch ab einem gewissen Punkt, die resultierenden Reads zeigten in ihren Chromatogrammen stark überlagerte Kurven.

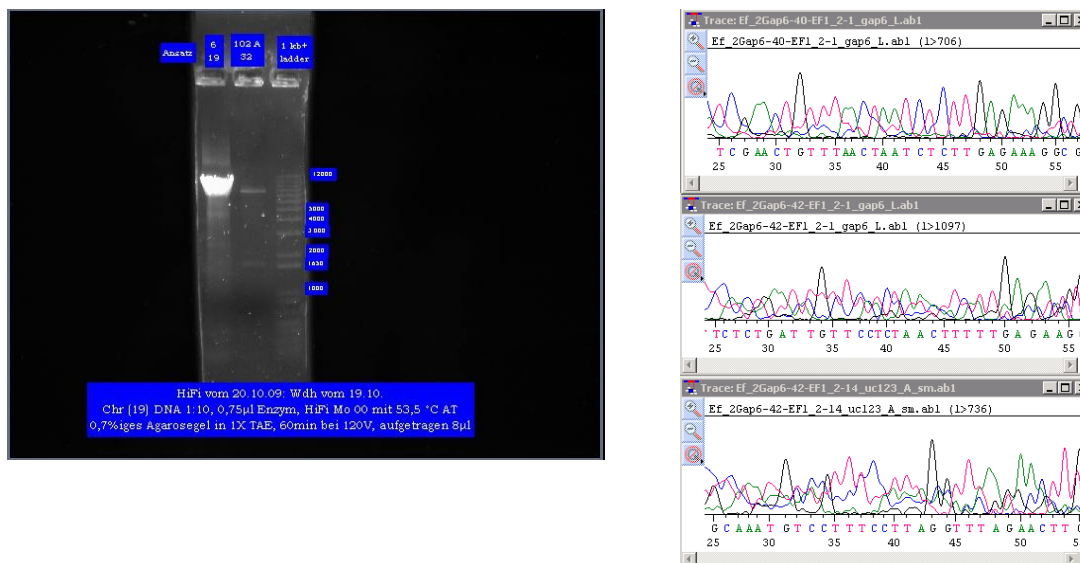


Abbildung 35: PCRs und Chromatogramme von Gap 6

4.6.1.2 Fosmide:

Für Gap 6 wurden Fosmide gefunden, welche die mutmaßlich neben dem Gap liegenden Contigs enthalten: Die Fosmid-Klone 1KA.D3, 2KB.C12, 2KB.F12, 2KB.F8 wurden unter diesen ausgewählt für die Sequenzierung. Zunächst wurde mit Primern, die in der Vektorsequenz binden, sequenziert, um die Ränder der Inserts zu sequenzieren. Hierbei zeigte sich, dass die Anordnung von Contig 6 und Contig 7 zueinander passte. Als nächstes wurde ebenso mit den für das Primerwalking entworfenen Primern sequenziert, wonach das gleiche Problem auftrat wie bei der Sequenzierung der PCRs: Die Reads waren aufgrund von überlagerten Chromatogrammen nicht mehr verwendbar.

Anhand des Restriktionsverdau-Musters mit BamH1 und des Mappings der mit den Vektor-ständigen Primern generierten Reads wurde zudem die Größe des verbleibenden Gaps kalkuliert.

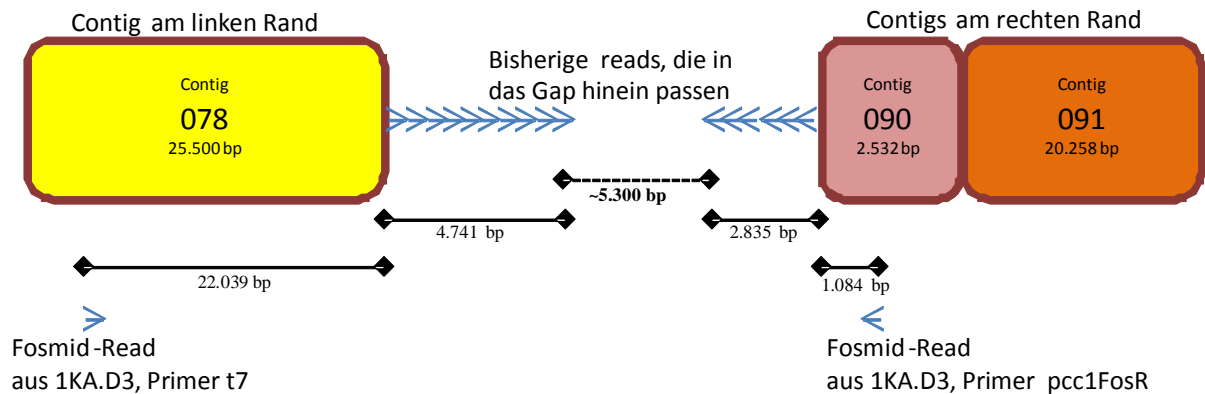


Abbildung 36: Bestimmung der Größe von Gap 6

Der Verdau des Fosmids 1KA.D3 ergab in etwa folgende Fragmente: 12.000 bp, 10.000 bp, 8.000 bp, 6.000 bp. Dies ergibt eine Gesamtlänge von etwa 36.000 bp. Zieht man hiervon die Längen der bisher bekannten Sequenzen, also die eigentlichen Contigs sowie die generierten Reads ab, erhält man als Rest 5.301 bp. Die Schätzung mittels der Elektrophorese ist natürlich nicht auf die Base genau, aber die Schätzung passt in etwa zu dem seitens der PCR generierten Produkt.

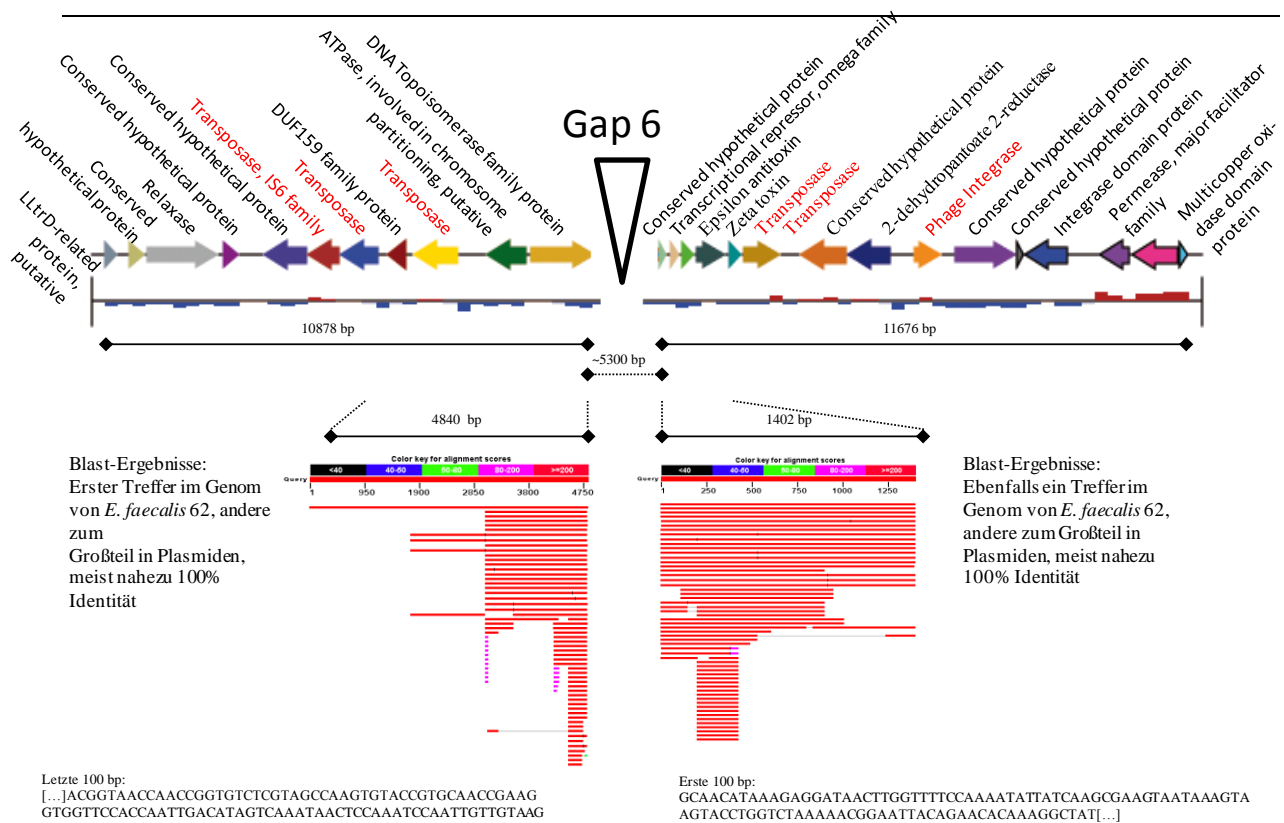
4.6.1.3 Annotation und Sequenzhomologien

Auf beiden Seiten des Gaps liegen mehrere Transposasen-Gene. Homologien zu den anderen Stämmen bestehen in diesem Bereich nur bedingt. Zwei CDS (conserved hypothetical protein und 2-dehydropantoate-2-reductase) sind jedoch in allen Referenzstämmen vorhanden, teilweise sogar mehrfach.

Auffallend sind weiterhin der zweite und der dritte CDS rechts neben dem Gap (transcriptional repressor, omega family und epsilon antitoxin). Diese kommen im Genom von *E. faecalis* 62 zweifach vor, das darauffolgende zeta toxin ist nur einfach vorhanden.

Eine Blast-Suche in der nucleotide-collection des NCBI ergibt, dass die Ränder des Gaps zum einen homolog zu Sequenzen innerhalb des Genoms von *E. faecalis* 62 sind. Zum anderen sind die Sequenzen homolog zu verschiedenen Plasmid-Sequenzen.

Eine Blast-Suche des ungemappten Contigs 084 ergibt, dass es ebenfalls im Genom von *E. faecalis* 62 passt.



Homologien zu den anderen Stämmen:

E. faecalis V583

E. faecalis OG1Rf

E. faecalis 62

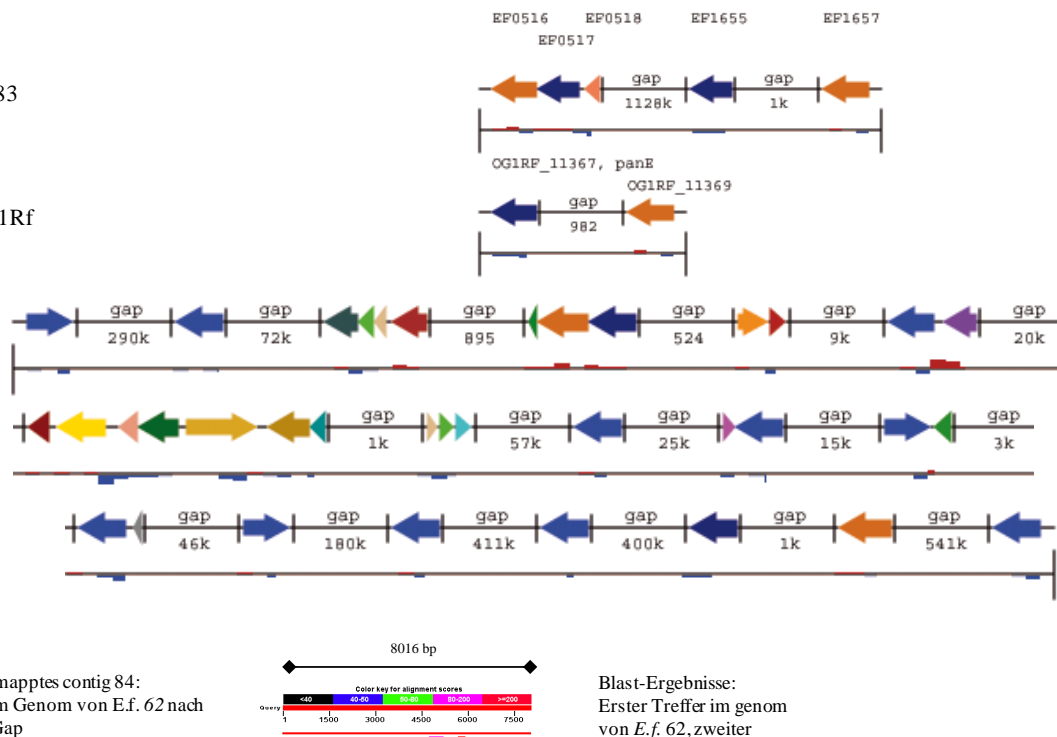


Abbildung 37: Übersicht über Gap 6

4.6.2 Übersicht über Gap 20

4.6.2.1 PCRs

Für Gap 20 konnte zunächst kein PCR-Produkt generiert werden.

Zum Zeitpunkt, als nur noch wenige Gaps vorhanden waren, wurden deshalb probeweise PCRs angesetzt, die Contigs an den Rändern der Gaps mit Contigs zu verbinden, die zu Beginn nicht an der Referenz *E. faecalis* V583 gemappt werden konnten.

Als zusätzliche Information stand ein Fosmid-Klon zur Verfügung, welcher nach Screening-PCR sowohl eine Sequenz vom linken Rand von Gap 20 enthielt als auch eine Sequenz von dem ungemappten Contig 102. Dementsprechend wurde eine PCR angesetzt und sequenziert und das ungemappte Contig 102 konnte daraufhin an den linken Rand von Gap 20 geknüpft werden.

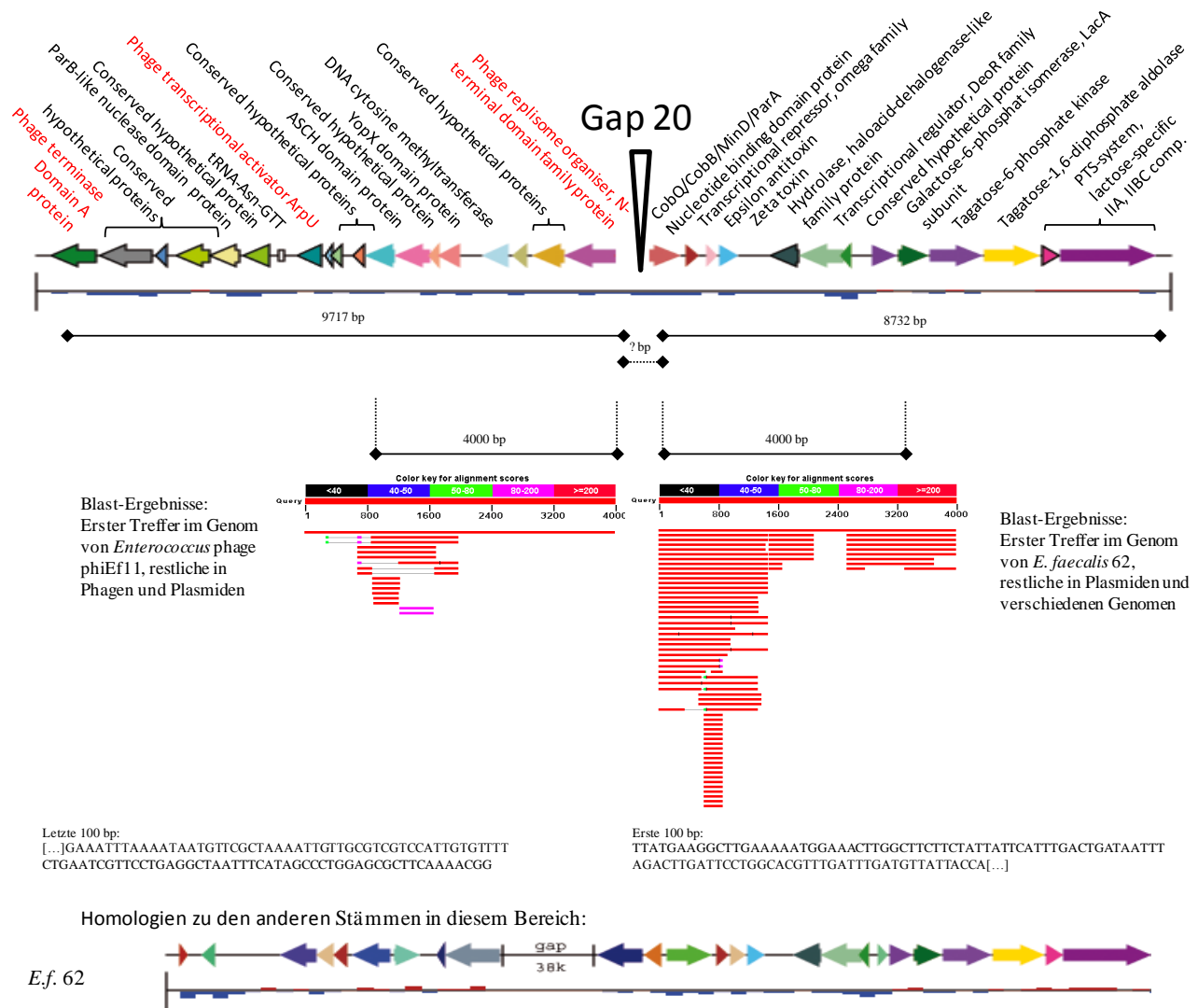
Ebenso konnte Contig 103, welches ursprünglich – mit eher geringer Homologie - in Gap 6 gemappt war, an den linken Rand von Gap 20 geknüpft werden.

Zwischen den beiden Contigs wurden PCRs angesetzt, die jedoch durchweg mehrfache Banden in der Elektrophorese zeigten. Dementsprechend konnten keine verwertbaren PCR-Produkte generiert werden.

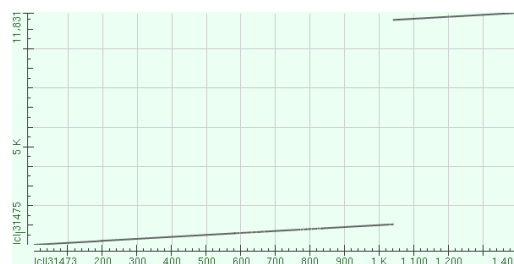
4.6.2.2 Fosmide

Um Contig 102 mit Contig 103 zu verbinden bzw. die Ausrichtung dieser Contigs zueinander zu überprüfen, wurde nun ein Fosmid gesucht, welches eben jene Contigs enthält. Mittels Screening-PCR wurden zu Fertigstellung dieser Arbeit jedoch nur drei Klone gefunden, die jedoch jeder nur jeweils eine der gesuchten Sequenzen enthalten

4.6.2.3 Annotation und Sequenzhomologien



Dot-matrix des Alignments der Contigs 123 und 103



Die Contigs, die jeweils den rechten Rand, beider Gaps bilden, sind stark homolog zueinander:

Abbildung 38: Übersicht über Gap 20

Die Sequenz links von Gap 20 beherbergt viele Phagengene. Sie ist weitgehend homolog zum Genom des *Enterococcus* phage phiEf11. Es lassen sich weitere Treffer in Phagen und Plasmiden finden.

Die rechte Seite hingegen ist homolog zu einem Teil des Genoms von *E. faecalis* 62, es bestehen weiterhin Ähnlichkeiten zu Plasmid-Sequenzen.

Ein Vergleich auf Gen-Ebene bestätigt die Homologie zu *E. faecalis* 62. Auffallend sind die CDS mit den Annotationen „transcriptional repressor, omega family“ und „epsilon antitoxin“. Homologe Gene bildeten den Rand von Gap 6. In *E. faecalis* 62 kommen sie ebenfalls doppelt vor.

Der rechte Rand beider Gaps scheint also sehr ähnlich zu sein. Die Contigs, die jeweils am Rand liegen, wurden mit Blast gegeneinander angeordnet. Die resultierende Dot-Matrix bestätigt die Homologie. In etwa die ersten 1.050 bp von Contig 123 entsprechen den ersten 1.050 bp von Contig 103, zusätzlich entsprechen die letzten ca. 350 bp einer gleich langen Sequenz etwas nach der Mitte des 17 kbp langen Contigs 103.

Aufgrund dieser Beobachtungen wurden die Contigs, welche zu Beginn des Assembly am Genom von *E. faecalis* V583 ausgerichtet wurden, nun am Genom von *E. faecalis* 62 testweise neu ausgerichtet (siehe Abbildung 39: Alternative Ausrichtung der Contigs, Seite 100).

5 Diskussion

5.1 Problemstellung

Bakterien gelten seit langem sowohl als Schädlinge als auch als Nutzbringer. Während manche Arten vor allem als Krankheitserreger bekannt sind und andere als größtenteils harmlos eingestuft werden, gibt es doch manche, die wie der Gott Janus zwei Gesichter zeigen können.

Enterokokken sind eine solche Gattung: Einerseits leben sie friedlich im Gastrointestinaltrakt des Menschen und beteiligen sich an der Verdauung der Nahrung und werden sogar teilweise bei der Nahrungsmittelproduktion selbst verwendet und als Heilmittel in Form von Probiotika verkauft (Foulquié Moreno et al. 2006). Andererseits sind sie sehr häufig an vergleichsweise harmlosen Infektionen wie Blasenentzündungen beteiligt und können sogar hin und wieder schwere, gar tödliche Entzündungen hervorrufen wie Pyelonephritis, Endokarditis und Sepsis (B. E. Murray 1990).

Es stellt sich die Frage, was macht bei Bakterien ein und derselben Gattung den Unterschied aus? Wie unterscheiden sich harmlose Nützlinge von gefährlichen Killern?

Im Rahmen des klinischen Umfelds ist hier die Resistenz gegenüber Antibiotika zu nennen. Gerade bei nosokomialen, also im Rahmen eines Klinikaufenthalts erworbenen Infektionen, sind Keime oft resistent gegenüber vielen Antibiotika, was eine Behandlung ungemein erschweren, in manchen Fällen sogar unmöglich machen kann.

Bekannt sind hier vor allem der MRSA (Methicillin-resistenter *Staphylococcus aureus*), und Enterobacteriaceae mit ESBL (erweitertes-Spektrum- β -Laktamase). Auch der eingangs erwähnte EHEC-Stamm, der den Ausbruch 2011 verursachte - er wurde später als Hybride aus EHEC (entero-hämorrhagischer *Escherichia coli*) und EAEC (entero-aggregativer *E. coli*) identifiziert und daher auch als EAHEC bezeichnet - trug eine solche ESBL (Mellmann et al. 2011; Brzuszkiewicz et al. 2011).

Gemeinsam ist beiden Bakteriengattungen, dass "herkömmliche" Vertreter von *S. aureus* und *E. coli* im Regelfall sensibel gegenüber den meisten Antibiotika sind.

Enterokokken sind jedoch von Natur aus resistent gegenüber den meisten Antibiotika, was die Therapieoptionen einschränkt. Die Entwicklung von zusätzlichen Resistenzen ist zum Glück bisher eher selten, kann dann jedoch richtig problematisch werden: Berüchtigt sind in den letzten Jahren so genannte VRE (Vancomycin-resistente Enterokokken) geworden, da Vancomycin bisher als Reserveantibiotikum für schwere Infektionen mit Kokken galt.

Der erste in den USA bekannt gewordene Stamm, *E. faecalis* V583 (Sahm et al. 1989), wurde in der Folge sequenziert, um den genetischen Aufbau zu untersuchen (Paulsen et al. 2003). Ein Ergebnis war, dass der Stamm von mobilen genetischen Elementen wie Phagen und Transposons durchsetzt war, die vermutlich zur Vermittlung der Antibiotika-Resistenz entschieden beitrugen.

Ebenso wurden eine Reihe von Pathogenitätsfaktoren publiziert, die für enterokokkale Infektionen verantwortlich gemacht werden, unter anderem Cytolysin, Aggregation Substance, Gelatinase, Collagen Adhesion Protein und Enterococcal Surface Protein (Jett et al. 1994; Rich et al. 1999; Domann et al. 2007).

Wie eingangs erwähnt sind jedoch nicht alle Enterokokken derart bedrohlich: Die meisten Menschen tragen viele Enterokokken im Darm ohne von ihnen vermutlich jemals Schaden zugefügt zu bekommen.

Ein anderer Sachverhalt ist jedoch die Einnahme von Enterokokken als Medikament, hier im Rahmen eines probiotischen Medikaments mit Namen Symbioflor 1. An ein Medikament wird die Anforderung der Unbedenklichkeit gestellt (§5, AMG, 2012).

Da in Symbioflor 1 lebende Enterokokken enthalten sind, sollte sichergestellt sein, dass diese möglichst nur das gute der beiden für Enterokokken mögliche Gesichter zeigen: Sie sollten den Wirt zwar - zumindest zeitweise - besiedeln können, sie sollten aber keine Infektion auslösen. Gerade eine Bakteriengattung, welcher unter Umständen nur schwer mit Antibiotika zu therapieren sein kann, sollte nicht leichtfertig an Patienten verabreicht werden - auch wenn die Gattung *Enterococcus* generell nur als fakultativ pathogen eingestuft wird.

Für die Ungefährlichkeit des Medikaments spricht, dass es nun seit ca. 60 Jahren eingenommen wird und bisher keine Infektionen dokumentiert wurden. Zudem entstammen die enthaltenen Enterokokken der Darmflora eines gesunden Menschen (Domann et al. 2007).

Um die Sicherheit des Stammes *E. faecalis* Symbioflor 1 zu bestätigen, wurde im Angesicht des mittlerweile bekannten Genoms von *E. faecalis* V583 die Entscheidung getroffen, den probiotischen Stamm auf genetischer Ebene zu untersuchen, um festzustellen, ob er in der Tat keine relevanten Pathogenitätsfaktoren enthält.

Zunächst wurde daher mittels PCRs und mittels Sequenzierung einer Shotgun-Plasmidbank überprüft, ob der Stamm relevante Pathogenitätsfaktoren enthält:

Demnach verfügte der Stamm genetisch über keine relevanten Pathogenitätsfaktoren außer über Aggregation Substance, Collagen Adhesion Protein und Resistenzgenen gegen oxidativen Stress. Experimentell konnte zudem Kapselbildung gezeigt werden. Für die anderen Pathogenitätsfaktoren waren die PCRs negativ und sie waren auch nicht im Assembly der Plasmidbank-Reads enthalten (Domann et al. 2007).

Problematisch war in dieser Hinsicht, dass eine negative PCR keinen absolut sicheren Ausschluss darstellt. Eine PCR ist darauf angewiesen, dass die Primer an ihre Zielsequenzen binden können. Unter Umständen könnte es passieren, dass Bindestellen für Primer mutiert sind, sodass eine Nachweis-PCR nicht mehr funktioniert, obwohl das Gen, was zwischen den Primern liegt, vollkommen intakt ist. Das Assembly der Plasmidbank-Reads erhöhte natürlich die Sicherheit der Aussage bezüglich der Abwesenheit der Pathogenitätsfaktoren, war mit 702 Contigs jedoch noch relativ lückenhaft.

Das Arzneimittelgesetz sieht zudem vor, dass Bestandteile von Medikamenten klar definiert sind (§22, AMG 2012). Ein lebendes Bakterium, welches seinen Zustand durch seinen Metabolismus ständig ändert, lässt sich am besten über sein Genom definieren. Mit Kenntnis der Genomsequenz ist prinzipiell auch die rRNA bekannt und damit die Proteine, und damit sind die Bestandteile relativ detailliert bestimmt.

Aus all diesen Gründen wurde beschlossen, das Genom von *E. faecalis* Symbioflor 1 möglichst komplett zu sequenzieren.

Aufgrund der hohen Anzahl an Contigs und dem damit verhältnismäßig großen Aufwand, diese mit PCRs zu schließen, wurde zunächst ein Lauf auf dem 454-Sequencer bei der Firma Roche veranlasst. Die so gewonnenen Reads wurden mit

den Reads aus der Plasmidbank zusammen mit dem Programm Newbler zu 69 Contigs assembliert. Von diesen konnten 45 am Referenzstamm *E. faecalis* V583 ausgerichtet werden. Daraufhin wurde begonnen, die Lücken zwischen diesen mit PCRs zu schließen. Für größere Lücken wurde zusätzlich eine Fosmidbank erstellt. Insgesamt konnte so das Genom von *E. faecalis* Symbioflor 1 sukzessive bis auf 2 Lücken geschlossen werden. Es resultierte ein Genom mit der Länge von 2.790.147 bp.

Die Genomsequenz wurde daraufhin in GenDB unter Zuhilfenahme von RAST, NCBI-Blast und Pfam annotiert (F. Meyer 2003; Aziz et al. 2008; Punta et al. 2012). Hierbei wurde ein halbautomatischer Ansatz gewählt:

Zunächst wurde die Sequenz von GenDB automatisch annotiert. Sodann wurde mit GeCo (C. T. Kuenne et al. 2007) eine Homologie-Matrix von *E. faecalis* Symbioflor 1 zu *E. faecalis* V583, dem zuerst vorhandenen Referenzgenom, erstellt (Paulsen et al. 2003). Alle GenDB-Annotationen, die mit einem homologen CDS des Referenzgenoms übereinstimmten, wurden übernommen. Die restlichen CDS wurden manuell in GenDB editiert.

Das fertig annotierte Genom wurde dann mit *E. faecalis* V583 sowie zwei weiteren mittlerweile erschienenen Genomen verglichen: *E. faecalis* OG1RF und *E. faecalis* 62.

5.2 Lückenschluss

Zwei Gaps sind bedauerlicherweise bei Fertigstellung dieser Arbeit immer noch offen.

Wie im Ergebnisteil beschrieben, stellen die Ränder dieser Gaps Sequenzen dar, in denen Transposasen und repetitive Sequenzen vorkommen. Die rechten Ränder beider Gaps sind zusätzlich sehr ähnlich zueinander. Der linke Rand von Gap 20 ist eine Region mit vielen Phagen-Genen.

Eine Region, die sich innerhalb des Genoms wiederholt, bietet einem Primer mehrere Bindungsstellen. So können zum einen bei einer PCR mehrere Produkte entstehen, zum anderen können Reads mit überlagerten Chromatogrammen entstehen.

Um einen guten Read zu erhalten, muss das Primer-Paar für die PCR daher möglichst spezifisch sein. Die Gelelektrophorese der PCR-Produkte für Gap 6 spricht dafür, dass ein solches, spezifisches Produkt generiert wurde.

Nun muss aber zusätzlich auch der Primer, mit dem sequenziert wird, zusätzlich spezifisch an eine einzige Stelle binden, zumindest an eine spezifische Stelle innerhalb des PCR-Produktes.

Dies scheint bei den Reads für Gap 6 nicht der Fall zu sein.

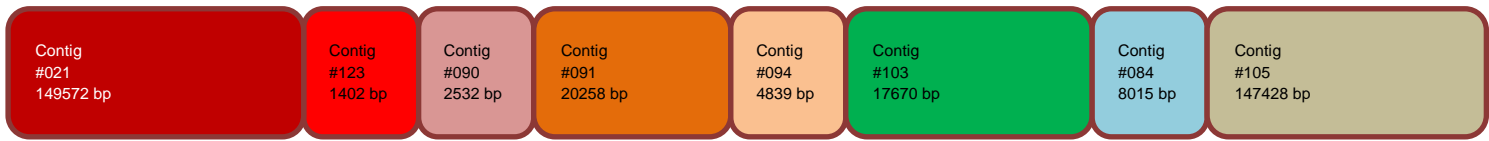
Dies kann verschiedene Gründe haben:

- Innerhalb des Gaps liegt eine Duplikatur von Sequenzen vor. Bindet der Primer an einer solchen Stelle, die sowohl innerhalb der bekannten Sequenz als auch innerhalb der unbekannten Sequenz vorkommt, werden verschiedene Amplifikate generiert, die zu überlagerten Reads führen.
- Innerhalb des Gaps liegt eine Sequenz vor, die eine inverse Kopie der Sequenz darstellt, welche den Primer bindet. Auch hier bindet der Primer ein zweites Mal und erzeugt ein unspezifisches Amplifikat. Diese Art eines inversen repeats spielt insbesondere bei der Funktion von Transposasen eine Rolle. Dementsprechend sind viele Gaps im ursprünglichen Assembly im Bereich von Transposasen gewesen. Zudem sind die entsprechenden Gaps oft schwerer als andere zu sequenzieren gewesen, denn es mussten öfter neue Primer generiert werden.

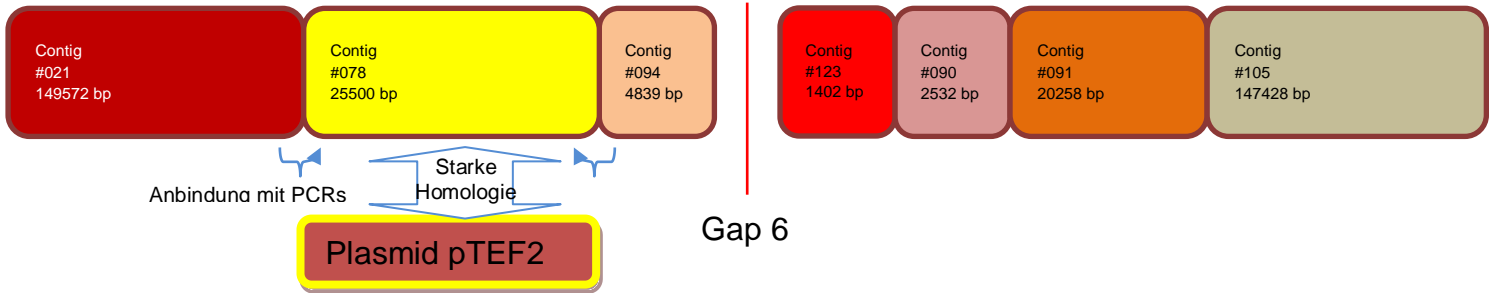
Für Gap 6 ist mit dem Genom von *E. faecalis* 62 nun noch ein Referenzgenom hinzugekommen, welches sich unter Umständen für die weitere Sequenzierung verwenden lässt. Lässt sich in der bisher bekannten Sequenz von *E. faecalis* Symbioflor 1 kein Primer in der Nähe des Gaps finden, der spezifisch bindet und verwertbare Reads in die unbekannte Sequenz hinein liefert, kann unter Umständen in der Referenz-Sequenz des verwandten Bakteriums ein Primer gefunden werden, der einen Read aus der unbekannten Sequenz heraus in die bekannte Sequenz hinein liefert.

Folgende Übersicht zeigt die ursprünglichen Contigs aus dem Hybrid-Assembly, wie sie anhand der Informationen aus *E. faecalis* 62 gemappt werden könnten.

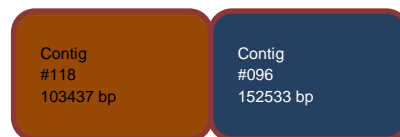
Contigs ausgerichtet am Genom von *E. faecalis* 62



Contigs im aktuellen Assembly, ursprünglich ausgerichtet am *E. faecalis* V583



Contigs ausgerichtet am Genom von *E. faecalis* 62



Contigs im aktuellen Assembly, ursprünglich ausgerichtet am *E. faecalis* V583

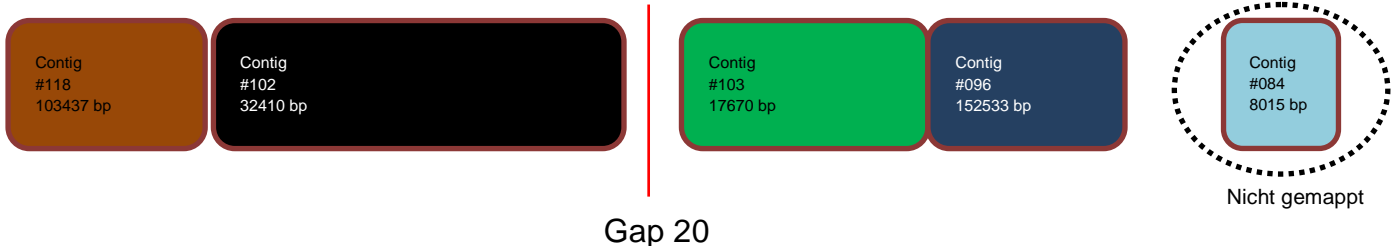


Abbildung 39: Alternative Ausrichtung der Contigs

Reihenfolge der randbildenden Contigs ausgerichtet am Referenzstamm *E. faecalis* 62 und im aktuellen *E. faecalis* Symbioflor 1-Assembly, welches auf einer Ausrichtung am Referenzstamm *E. faecalis* V583 basiert. Länge der Blöcke nicht streng maßstabsgetreu. Erstellt nach einem Alignment in MAUVE

Es fällt auf, dass zwei große Contigs (078 und 102) im Genom von *E. faecalis* 62 keine Entsprechung finden. Contig 078 stellt den linken Rand des Gap 6 dar, Contig 102 den linken Rand von Gap 20. Zudem sind die Contigs 094, 103 und 123 im Assembly von *E. faecalis* Symbioflor 1 an andere Stellen gemappt als die entsprechenden Sequenzen im Genom von *E. faecalis* 62.

Es stellt sich die Frage, ob diese Veränderungen das Resultat von genomischen Rearrangements sind, oder ob sie auf fehlerhaft zugeordneten Sequenzen beruhen.

Die PCR-Reads, die die entsprechenden Contigs verbinden, sprechen dafür, dass Rearrangements vorliegen, allerdings ist dies nicht mit absoluter Sicherheit zu sagen, wenn der gesamte Bereich, in dem eine PCR liegt, mehrfach im Genom vorliegt.

Die Zuordnung der Contigs zueinander könnte mit längeren PCRs erneut überprüft werden, deren Primerpaare sehr weit entfernt vom ehemaligen Gap in Regionen liegen, die außerhalb der wiederholten Struktur liegen.

Zudem könnte versucht werden, die Contigs mit entsprechenden PCRs analog zur Referenz *E. faecalis* 62 zu verbinden.

Für Contig 103 gilt ebenso, dass die Positionierung innerhalb von Gap 20 überprüft werden sollte. In diesem Zusammenhang sollte versucht werden, das ungemappte Contig 084 zwischen die Contigs 103 und 105 zu legen.

Für Gap 20 sollte jedoch zunächst versucht werden, Primer zu entwerfen, die weit entfernt von Rändern des Gaps liegen, insbesondere die Region, die in Gap 6 und Gap 20 am rechten Rand vorkommt, sollte für die Primerfindung gemieden und somit mittels PCR übersprungen werden.

Weiterhin sollten die drei Fosmide, die Rand-Sequenzen des Gaps enthalten, mit den Vektorprimern und mit den Primern der enthaltenen Contigs sequenziert werden. So kann zumindest Aufschluss darüber gewonnen werden, welche Contigs in der Nähe liegen, ggfs. kann Primerwalking durchgeführt werden.

Eine weitere Möglichkeit um zu verhindern, dass durch doppelt bindende Primer die Sequenzierreaktion gestört wird, wäre folgende: Ein Fosmid-Insert, welches das entsprechende Gap enthält, oder ein PCR-Produkt, falls vorhanden, wird noch einmal in kleine Stücke (500 - 1.000 bp) verdaut oder geschert. Diese werden dann in einen Plasmidvektor subkloniert, welcher in ein Bakterium transformiert wird. Die so gewonnene Plasmid-Bibliothek wird daraufhin mit Vektor-seitigen Primern sequenziert. Auf diese Weise würde umgangen, dass Sequenzier-Primer in repetitiver Sequenz binden.

Die resultierenden Reads werden in einem Assembly zur Gap-Sequenz zusammengesetzt und dann mit den entsprechenden Contigs verknüpft.

Eine andere Möglichkeit, repetitive Sequenzen von Fosmid-Inserts oder langen PCRs zu überbrücken, wäre ein sequentieller Abbau: Mit einer S1-Nuklease werden

vom Ende des Inserts oder des PCR-Amplifikates Nukleotide abgespalten. Dieser Prozess wird zu verschiedenen Zeitpunkten gestoppt, sodass unterschiedlich lange Fragmente resultieren. Diese werden wiederum in einen Plasmid-Vektor kloniert. Wenn nun dieses Plasmid-Insert mit einem vektorständigen Primer sequenziert wird, kann in jede Stelle des Fosmid-Inserts oder des PCR-Produktes hinein sequenziert werden, solange ein entsprechend verdautes Fragment zur Verfügung steht.

Die meisten Contigs in der Nähe der Gaps zeigen Homologien zu Plasmiden. Insbesondere Contig 078 zeigt starke Homologien zu Plasmid pTEF2, einem Plasmid von *E. faecalis* V583, auf dem auch eine Kopie der Aggregation Substance vorliegt. Es stellte sich die Frage, ob diese Sequenz nicht zum Chromosom, sondern zu einem Plasmid gehört.

Bisher konnte kein Plasmid bei *E. faecalis* Symbioflor 1 nachgewiesen werden. Wiederholte Experimente mittels PFGE, die auch Plasmide mit geringen Kopien pro Bakterienzelle detektieren sollten, ergaben keinen Hinweis auf ein unter den angewandten Wachstumsbedingungen frei vorhandenes Plasmid. (Domann, persönliche Kommunikation, Domann et al. 2007).

Es ist jedoch nicht ungewöhnlich, dass Plasmid-Sequenzen in ein Genom integriert werden. Auch stellen sie oft Gegenstand von genomischen Rearrangements dar, sodass das widersprüchliche Mapping der Contigs anhand von *E. faecalis* 62 nicht unbedingt auf eine falsche Anordnung der Contigs in *E. faecalis* Symbioflor 1 hinweist. Es ist auch möglich, dass ein Plasmid unter bestimmten Wachstumsbedingungen episomal vorliegt.

Die Anbindung des Contigs 078 an die Nachbar-Contigs wurde mittels PCRs überprüft. und bestätigt die Lage zwischen den Nachbar-Contigs.

Theoretisch könnte das Contig 078 während des *de novo* Assemblies fehlerhaft zusammengesetzt worden sein. Dazu muss gesagt werden, dass an beiden Enden von Contig 078 Transposasen-Sequenzen vorliegen. Diese stören durch ihre repetitiven Elemente das Assembly. Die Nahtstellen, an denen Plasmid-homologe Sequenzen in nicht homologe Sequenzen übergehen, wurden daraufhin im Assembly

Viewer von Newbler geprüft. Es ergab sich kein deutlicher Hinweis auf ein Fehl-Assembly.

Eine andere Erklärung wäre, dass *E. faecalis* Symbioflor 1 ein Plasmid besitzt, welches mit den herkömmlichen Isolierungs-Methoden nicht erfasst wurde, jedoch bei der Isolierung der chromosomalen DNA zumindest in Teilen mit isoliert wurde, und dass die Contigs mit hoher Homologie zu Plasmiden in der Tat nicht zum Chromosom gehören. Dem widerspricht, dass zu manchen dieser Contigs entsprechende Sequenzen im Chromosom von *E. faecalis* 62 gefunden wurden. Eine Integration von Plasmid-Sequenzen in *Enterococcus*-Genome ist also nicht unbedingt ungewöhnlich.

5.3 Fazit für die Strategie zum Lückenschluss prokaryontischer Genome

Nach der *de novo*-Assemblerung der Reads aus dem 454-Pyrosequenzierungs-Verfahren mit den Sanger-Reads aus der Plasmid-Bibliothek blieben Gaps übrig, die des öfteren im Bereich von Transposasen lagen. Ebenso lagen alle 4 rRNA-Kassetten in Gaps. Dies legt die Vermutung nahe, dass die Gaps nicht rein zufällig verteilt sind. Vielmehr liegen sie häufig in Bereichen mit Sequenzen, die in Kopien vorliegen, entweder relativ nahe beieinander oder wie im Fall der rRNA-Kassetten an weit entfernten Loci. Dies lässt sich dadurch erklären, dass ein Assemblierungs-Algorithmus die Reads aufgrund von Sequenz-Übereinstimmungen aneinanderreicht. Wenn nun aber eine Sequenz doppelt im Genom enthalten ist, können Reads aus diesen Sequenzen nur dann eindeutig zugeordnet werden, wenn sie weit genug aus der doppelten Sequenz herausragen um zu einem der Loci zugeordnet zu werden. Liegt ein Read komplett innerhalb doppelt vorhandener Sequenz, kann er nicht zugeordnet werden.

Konsequenterweise wird also an dieser Stelle ein Gap offen gelassen.

Um dies zu vermeiden, empfiehlt es sich, möglichst lange Reads zu generieren. Je länger die Reads, desto sicherer können sie zugeordnet werden.

Die Qualität und Anzahl der Reads spielen ebenso eine Rolle. Im Idealfall, wenn jede Base jedes Reads 100%ig sicher wäre, könnten Reads an Stellen, die sich nur durch eine Base unterscheiden, immer noch richtig zugeordnet werden. Je mehr Fehler in den Reads sind, desto schlechter kann ein Assembly-Programm Reads eindeutig einem Locus zuordnen, und konsequenterweise muss es mehr Gaps offen lassen. Über die Anzahl der Reads können zumindest die zufälligen Fehler in den Basen bedingt ausgeglichen werden. Dennoch bleibt die Länge der Reads ein limitierender Faktor, selbst wenn sehr viele sehr hochwertige Reads vorhanden sind, wie das obige Beispiel zeigt.

Eine andere Möglichkeit sind paired end (oder mate pair) Reads: Sie sind zum Beispiel im Rahmen der Plasmid-Bibliothek verwendet worden: Wenn jedes Plasmid-Insert von beiden Seiten sequenziert wird, sind diese beiden Reads zueinander entgegengesetzt anzuordnen. Dies bedeutet eine Zusatzinformation für den Assemblierungs-Algorithmus, mit dem dieser unter Umständen Reads von mehrfach vorhandenen Sequenzen eindeutig zuordnen kann. Mittlerweile sind solche paired-end-Sequenzierungen auch für 454-Pyrosequenzierungs-Läufe verfügbar, und sollten aus dieser Überlegung heraus bevorzugt Verwendung finden.

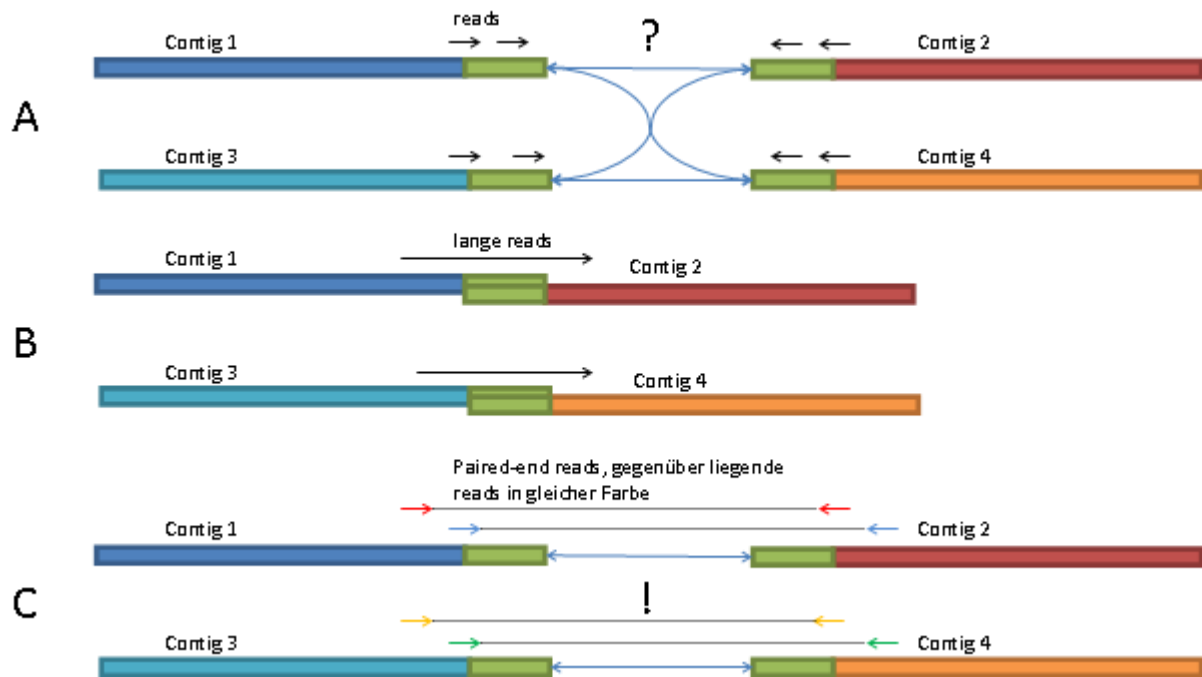


Abbildung 40: Das Problem, repeats zu assemblieren

Die zueinander homologen Sequenzen sind in grün gezeichnet.

A zeigt den Problemfall: Die Contigs können auf unterschiedliche Weise verbunden werden. Selbst wenn sehr viele qualitativ einwandfreie Reads vorhanden sind, kann keine Entscheidung getroffen werden, welche Contigs zueinander gehören, weil die Reads zu kurz sind.

B zeigt eine mögliche Lösung: Lange Reads, die in die nicht homologen Sequenzen hineinragen, überbrücken das Gap komplett.

C zeigt eine zweite Lösung: Mit gepaarten Reads können Contigs einander zugeordnet werden

Es empfiehlt sich weiterhin, die Primer, die für den Lückenschluss verwendet werden, primär etwas weiter entfernt von den Rändern der Gaps zu wählen: Die Ränder enthalten oft repetitive Sequenzen, die für das Setzen von PCR-Primern vermieden werden sollten.

Zum Beispiel könnte die Standard-Entfernung der Primer zum Gap für die Software minimap auf 1.000 bp erhöht werden. Ein hiermit generierter Read würde zwar weniger weit in das Gap hineinragen, das Risiko für eine unspezifische PCR würde jedoch verringert. Zum Sequenzieren dieser PCRs könnten weiterhin Primer verwendet werden, die bis 150 bp nahe am Gap liegen, in der Hoffnung, dass die Primer zumindest innerhalb des PCR-Produkts spezifisch binden.

Aus derselben Überlegung heraus sollte überlegt werden, die minimale Größe der verwendeten Contigs von 500 bp auf 1.000 bp zu erhöhen.

Die kleineren Contigs sind oft das Ergebnis von repeat-reichen Strukturen, in deren Umgebung oft keine eindeutig sequenzierbaren PCRs generiert werden konnten.

Da es im Rahmen der Arbeit möglich war, PCRs bis etwa 10.000 bp zu generieren, und die Reads aus guten PCRs bis knapp über 1.000 bp lang waren, erscheint es sinnvoll, die kleineren Contigs zunächst wegzulassen, und PCRs von größeren benachbarten Contigs aus zu entwerfen, die die kleinen Contigs überspannen und deren Primer in Regionen ohne Repeats liegen.

5.4 Putative Pathogenitätsfaktoren

Wie schon eingangs erwähnt, sollte bei Verwendung eines Bakterienstammes am Menschen eine Bedrohung durch mögliche Pathogenitätsfaktoren soweit wie möglich ausgeschlossen werden.

Zunächst hat die Genomsequenzierung gezeigt, dass einige Faktoren vorhanden sind: *E.faecalis* Symbioflor 1 verfügt über Aggregation Substance, Collagen Adhesion Protein und über die Gene, die er zum Widerstand gegen oxidativen Stress benötigt. Weiterhin konnte gezeigt werden, dass der Stamm in der Tat nicht über Cytolysin, Enterococcal Surface Protein, Gelatinase, Hyaluronidase oder Peptid-Antibiotikum AS-48 verfügt.

Die Aggregation Substance ist wichtig für die Adhäsion von Enterokokken aneinander zum Austausch von Plasmiden wodurch potentiell pathogenen Stämmen die Aufnahme von z.B. Antibiotika-Resistenzen vermittelt wird (Dunny 1990). Außerdem vermittelt Aggregation Substance die Adhäsion an eukaryontische Zellen, dies spielt möglicherweise bei der Ausbildung von Harnwegsinfekten eine Rolle (Kreft et al. 1992). Auch bei einer Endokarditis bilden Stämme mit Aggregation Substance dichtere Vegetationen auf dem Endokard. Eine hohe Mortalität bei einer Endokarditis scheint jedoch nur in Kombination mit Cytolysin erreicht zu werden (Chow et al. 1993).

Die Gene für die gewebeschädigenden Lysine Gelatinase, Cytolysin und Hyaluronidase konnten in *E. faecalis* Symbioflor 1 nicht gefunden werden. Der Verlust der Gelatinase ist möglicherweise auf die Insertion einer Transposase zurückzuführen.

Damit ist davon auszugehen, dass der Keim vermutlich schlechter menschliches Gewebe infiltrieren kann als vergleichbare Enterokokken mit diesen Genen.

Aggregation Substance erscheint in diesem Licht als Faktor, der Enterokokken die Kolonisation ermöglicht, während Faktoren wie Cytolysin, Gelatinase und Hyaluronidase zur Zerstörung von Wirtszellen und Gewebe führen. Für einen probiotischen Stamm ist es wünschenswert, dass er kolonisiert ohne zu zerstören, und diese Kombination scheint mit der Anwesenheit von Aggregation Substance und Abwesenheit von Gewebs-schädigenden Enzymen wie Cytolysin, Gelatinase und Hyaluronidase vorzuliegen.

Es bleibt noch die Frage, ob *E. faecalis* Symbioflor 1 Aggregation Substance tatsächlich exprimiert. Im Vergleich zu *E. faecalis* V583, der zwei Kopien des Gens hat, liegt im probiotischen Stamm nur ein Gen vor.

Das Collagen Adhesion Protein spielt eine Rolle bei Harnwegsinfekten und Endokarditiden, indem es den Enterokokken ermöglicht, sich an die extrazelluläre Matrix des Wirtsorganismus zu heften (Nallapareddy et al. 2011). Das Gen für dieses Protein ist in allen vier untersuchten Stämmen vorhanden, jedoch ist es sowohl in *E. faecalis* Symbioflor 1, als auch in *E. faecalis* 62 verkürzt und damit vermutlich funktionell inaktiv.

Dies beeinträchtigt möglicherweise diese beiden Stämme dabei, Harnwegsinfekte oder Endokarditiden auszulösen.

Zum Ausbilden eines Biofilms, wie er bei vielen klinischen Isolaten gefunden wird, ist das Enterococcus Surface Protein erforderlich (Tendolkar et al. 2004; Toledo-Arana et al. 2001).

Von den untersuchten Stämmen enthielt nur *E. faecalis* 62 das entsprechende Gen. Dies passt zu der Beobachtung, dass der Stamm *E. faecalis* Symbioflor 1 experimentell keinen Biofilm bilden konnte (Domann et al. 2007).

Für die Abwehr gegen oxidativen Stress wurde seinerzeit festgestellt, dass *E. faecalis* Symbioflor 1 über die entsprechenden Genausstattung verfügt (Domann et al. 2007). In der komparativen Analyse mit den anderen Stämmen fiel nun auf, dass die Gene für Katalase sowie für eine Kopie der Thioredoxin-Reduktase verkürzt sind.

Das Enzym Katalase ermöglicht Bakterien, toxisches Wasserstoff-Peroxid (H_2O_2) zu H_2O und Sauerstoff umzusetzen. Unter anaeroben Bedingungen, wie sie im Darm herrschen, entsteht relativ wenig Wasserstoff-Peroxid. Unter aeroben Bedingungen produziert der Stoffwechsel jedoch als Nebenprodukt Superoxid und Wasserstoff-Peroxid. Auch die Atmungskette, die den Keim unter aeroben Bedingungen mit Energie versorgt, verursacht vermutlich die Entstehung von Sauerstoff-Radikalen und letztlich Wasserstoff-Peroxid. Es konnte gezeigt werden, dass *E. faecalis* über eine Katalase (*katA*) verfügt, die dem Keim ein Überleben in Gegenwart von Wasserstoff-Peroxid ermöglicht. Diese Katalase benötigt jedoch Häm als prosthetische Gruppe. Dieses stellt der Keim nicht selbst her, kann es aber aus dem Wachstumsmedium entnehmen (Pritchard & Wimpenny 1978; Ritchey & Seeley 1974; Frankenberg et al. 2002).

In der klinischen Diagnostik wird 3%iges Wasserstoff-Peroxid benutzt, um auf Agarplatten zwischen Staphylokokken, Streptokokken und Mikrokokken unterscheiden zu können. Staphylokokken, als Katalase-positive Keime, setzen das H_2O_2 um, der entstehende Sauerstoff ist als Bläschenbildung im Agar sichtbar. Streptokokken sind in der Regel Katalase-negativ. Enterokokken, die historisch zu den Streptokokken gehören, gelten ebenfalls als Katalase-negativ, bei ihnen wird in der Regel keine oder nur schwache Bläschenbildung durch eine sogenannte Pseudokatalase beobachtet, die bestenfalls nach frischer Isolierung auftritt und dann oft nach serieller Kultivierung verschwindet (Winn, et al., 2006).

Es ist möglich, dass die Enterokokken beim Wachstum auf den in der Diagnostik benutzen Platten das benötigte Häm nicht zur Verfügung haben. Werden Blutplatten verwendet, wären Erythrozyten und damit Häm vorhanden, aber dazu müssten die Enterokokken die Erythrozyten lysieren um an das Häm zu gelangen. Es ist auch möglich, dass die enterokokkale Katalase unter diesen Wachstumsbedingungen

nicht ausreichend exprimiert wird, um mit bloßem Auge die Bläschenbildung bei Anwendung des H_2O_2 -Tests sehen zu können.

Es ist plausibel, dass die Produktion von Wasserstoff-Peroxid und Sauerstoff-Radikalen sowie die Verteidigung gegen ebensolche Stoffe den Enterokokken ermöglicht, während einer Infektion in menschlichem Gewebe und im Blut zu überleben, vor allem da auch Leukozyten diese als Angriffsmechanismen benutzen (Frankenberg et al. 2002).

Der probiotische Stamm *E. faecalis* Symbioflor 1 verfügt jedoch offensichtlich nicht über eine funktionierende Katalase. Im Katalase-Test ist der Stamm wie erwartet negativ. Die entsprechende CDS zeigt eine Punktmutation mit einem daraus resultierenden Stopp-Codon und daher ein fragmentiertes Protein. Es ist anzunehmen, dass der Keim dadurch anfälliger gegenüber neutrophilen Leukozyten ist. Zudem dürfte er gegenüber den Sauerstoff-Radikalen, die bei Wachstum in lebendem Gewebe unter aeroben Bedingungen vermehrt anfallen, sensibler sein als Enterokokken mit intakter Katalase.

Sauerstoff-Radikale sind jedoch nicht nur im Kampf mit Leukozyten bedeutsam. Milchsäurebakterien produzieren auch während der oxidativen Energiegewinnung aus Lactat, Pyruvat und Glycerol ebenfalls H_2O_2 . Wenn dieses akkumuliert, ist es toxisch für die Bakterien, indem es die DNA, Proteine und die Zellmembran schädigt. Es scheint, als ob verschiedene Stämme des *Enterococcus faecalis* die Möglichkeiten des oxidativen und des anaeroben Abbaus von Glycerol in unterschiedlichem Maße nutzen, auch wenn sie prinzipiell beide Wege zur Verfügung haben. Es hat sich jedoch gezeigt, dass *Enterococcus*-Deletionsmutanten, denen die Möglichkeit genommen wurde, Glycerol oxidativ abzubauen, in Gegenwart von Glycerol besser wuchsen als der Wildtyp. Offensichtlich ist das vom Keim selbst produzierte H_2O_2 für sein eigenes Wachstum hinderlich. Im Gegensatz dazu wuchsen die *Enterococci* besonders schnell, wenn sie den oxidativen Weg zur Verfügung hatten und zusätzlich mit Katalase im Medium versorgt wurden (Bizzini et al. 2010).

Eine weitere Bedrohung durch oxidativen Stress verursachen Antibiotika. Es konnte gezeigt werden, dass bakterizide Antibiotika wie Penicillin und Vancomycin

zumindest einen Teil ihrer Wirkung der Entstehung von Superoxidanionen (O_2^-) verdanken. Diese negativ geladenen Sauerstoff-Moleküle werden von der Superoxid-Dismutase zu H_2O_2 verstoffwechselt. Dieses kann dann von der Katalase entgiftet werden. Insbesondere Funktionseinbußen an der Superoxid-Dismutase erhöhen die Empfindlichkeit der Enterokokken gegenüber jenen Antibiotika. Die Katalase, welche erst den zweiten Schritt der Entwicklung katalysiert, scheint hier weniger wichtig zu sein, Funktionseinbußen machen sich weniger bemerkbar (Bizzini et al. 2009).

Um zu testen, inwieweit solche Überlegungen zutreffen, können auf Basis der vorhandenen Genomsequenz gezielt Deletionsmutanten von *E. faecalis* Symbioflor 1 hergestellt werden, deren Gene für Katalase oder Thioredoxin-Oxidasen deletiert sind. Diese können dann unter verschiedenen Wachstumsbedingungen auf ihre Resistenz gegenüber oxidativem Stress getestet werden.

Relevant könnte dies auch unter dem Aspekt sein, dass Sauerstoff-Radikale auch für eukaryontische Zellen schädlich sind. Die Schädigung der Darmschleimhaut mit Radikalen wurde als ein potentieller Faktor für die Entstehung des colorektalen Karzinoms und anderer, entzündlicher Darmerkrankungen diskutiert (Y. J. Kim et al. 2012). Es könnte also durchaus eine Rolle spielen, wie viel oxidativen Stress die eigene Darmflora verursacht oder ob sie diesen sogar mindert.

Für einen probiotischen Stamm wie *E. faecalis* Symbioflor 1 wäre in diesem Zusammenhang zunächst interessant, welchen Stoffwechselweg er bevorzugt: Jenen, der vermehrt Radikale freisetzt, oder den, bei dem dies nicht der Fall ist. Die vermutlich nicht funktionierende Katalase ist also durchaus aus zwei Blickwinkeln zu betrachten: Während sie auf der einen Seite dem Keim die Abwehr gegen menschliche Abwehrzellen erschwert, führt sie unter Umständen zur vermehrten Ansammlung von schädlichen Radikalen.

Es konnte gezeigt werden, dass Stämme mit Cytolysin vermehrt im Blut auftauchen und proliferieren. Unter der Annahme, dass die Enterokokken mit Cytolysin menschliche Erythrozyten lysieren, um deren Häm für ihre Katalase (und andere Enzyme) zu verwenden, fällt auf, dass *E. faecalis* Symbioflor 1 auch diese Möglichkeit fehlt, da er nicht über Cytolysin verfügt. Ein Überleben von Symbioflor 1

im menschlichen Blutstrom erscheint also insgesamt deutlich erschwert (Pillar & Gilmore 2004).

Der Locus, welcher *E. faecalis* V583 die namensgebende Vancomycin-Resistenz verleiht, liegt in einer Pathogenitätsinsel, deren Ausmaß und Inhalt zwischen den Stämmen sehr unterschiedlich ist. Diese Beobachtung deckt sich mit früheren Analysen des Locus (Dahl et al. 1999). Das *vanB*-Operon selbst besitzt von den in dieser Arbeit untersuchten nur jener Stamm, während zumindest die Stämme *E. faecalis* 62 und *E. faecalis* OG1RF noch Teile der umgebenden Region besitzen. In *E. faecalis* Symbioflor 1 fehlt dieser Locus im Vergleich zu den anderen Stämmen jedoch komplett.

Somit kann auch genetisch betätigt werden, dass der Stamm nicht Vancomycin-resistent ist.

Resistenzmechanismen werden zum Teil über mobile genetische Elemente übertragen, was insbesondere auch für das *vanB*-Operon gilt (Quintiliani & Courvalin 1994).

E. faecalis Symbioflor 1 zeigt, dass auch ein Stamm mit relativ vielen chromosomalen mobilen genetischen Elementen nicht unbedingt auch viele Pathogenitätsfaktoren haben muss.

Weiterhin wurde geschlussfolgert, dass eine Resistenz gegenüber fremder DNA eine wünschenswerte Fähigkeit von apathogenen Keimen ist, um so die Aufnahme von Resistenz-Genen zu verhindern (Palmer & Gilmore 2010). CRISPR-Systeme sind solche Systeme (Makarova et al. 2006; Barrangou et al. 2007).

Es fällt auf, dass *E. faecalis* Symbioflor 1 rein von der Anzahl nach *E. faecalis* OG1RF die beste Ausstattung an CRISPR-Loci enthält. Im Gegensatz dazu hat der Vancomycin-resistente *E. faecalis* V583 nahezu keinen. Die Assoziation kann in diesem Fall leider nicht uneingeschränkt übernommen werden, da *E. faecalis* Symbioflor 1 kein Enzym für ein funktionierendes CRISPR-System aufweist. Damit ist fragwürdig, ob die entsprechenden Loci überhaupt eine Funktion haben - es sei denn, ein bisher unbekanntes Protein würde die Funktion bekannter CRISPR-Proteine übernehmen.

Zudem besitzt *E. faecalis* Symbioflor 1 im Gegensatz zu *E. faecalis* OG1RF, der fast gar keine mobilen genetischen Elemente aufweist, einige Prophagen und Transposasen, was gegen ein funktionierendes CRISPR-System spricht.

Zusammenfassend besitzt *E. faecalis* Symbioflor 1 Faktoren, die die Besiedlung eines Wirtsorganismus begünstigen, aber die Faktoren zur Infektion, Invasion und Schädigung von Gewebe fehlen ihm. Insbesondere die Gene für Cytolysin, Gelatinase und Enterococcal Surface Protein scheinen klinisch bedeutsame Isolate von harmlosen Stuhlkeimen zu unterscheiden, und keiner dieser Faktoren ist in *E. faecalis* Symbioflor 1 enthalten. Einzig Aggregation Substance ist ein Genprodukt, welches auch vermehrt in pathogenen Isolaten gefunden wurde (Semedo et al. 2003). Diese und das ebenfalls vorhandene Collagen Adhesion Protein sowie die Gene, die Widerstand gegen oxidativen Stress vermitteln, sind jedoch vermutlich für die Kolonisation im Darm nützliche Faktoren, ohne direkt schädlich zu wirken (Domann et al. 2007).

Da der Stamm *E. faecalis* Symbioflor 1 wesentliche Pathogenitätsfaktoren nicht hat, ist unwahrscheinlich, dass er eine Krankheit auslösen kann. Er wurde aufgrund dessen von der zentralen Kommission für biologische Sicherheit von Risikostufe 2 gemäß Gentechnik-Sicherheitsverordnung auf Risikostufe 1 heruntergestuft.

Er gilt damit als harmlos und kann bedenkenlos eingenommen werden.

5.5 Überleben der Magen-Darm-Passage

Um als Probiotikum eine Wirkung in der Darmflora entfalten zu können, muss ein oral appliziertes probiotisches Bakterium lebend dort ankommen. Auf diesem Weg sind vor allem die Magensäure und Gallensäften Faktoren, die auf Bakterien äußerst schädlich wirken. Christoffersen et al. untersuchten *in vitro* die Empfindlichkeit verschiedener *E. faecalis*-Stämme gegenüber Magensaft und Duodenalsaft. Hierbei zeigte sich, dass der Magensaft, hauptsächlich aufgrund seiner Azidität, sowohl *E. faecalis* Symbioflor 1 als auch die Stämme *E. faecalis* 62 und *E. faecalis* V583 gleichermaßen abtötet. Gegenüber Duodenalsaft zeigte *E. faecalis* Symbioflor 1 jedoch eine erhöhte Resistenz (Christoffersen et al. 2012).

Bøhle et al untersuchten *E. faecalis* V583 auf Proteom-Ebene, um Gene zu finden, die das Überleben in Gegenwart von Gallensäure ermöglichen (Bøhle et al. 2010).

Es zeigt sich, dass *E. faecalis* Symbioflor 1 über alle bis auf fünf dieser Gene verfügt – vier Proteine mit unbekannter Funktion und ein Mannose-spezifisches Transportsystem fehlen.

Diese Ergebnisse lassen vermuten, dass *E. faecalis* Symbioflor 1 relativ gut im Dünndarm in Gegenwart von Duodenalsaft und den dort enthaltenen Gallensäuren überleben kann, jedoch die Magenpassage aufgrund des sauren pH-Wertes nicht überlebt.

Andererseits besitzt der Stamm Faktoren wie Aggregation Substance, Collagen Adhesion Protein und Resistenz gegenüber oxidativem Stress, die ein Überleben im Darm begünstigen - zudem wurde der Keim ursprünglich aus dem Stuhl eines gesunden Menschen isoliert. (Domann et al. 2007).

Es ist also ohne weiteres nicht abzuleiten, ob oral eingenommene *E. faecalis* Symbioflor 1-Bakterien die Magen-Darm-Passage überleben und sich im Darm ansiedeln.

Hierzu wäre ein direkter Nachweis des Stammes aus dem Stuhl einer Versuchsperson erforderlich. Da die meisten Menschen jedoch eigene *E. faecalis*-Stämme in ihrer Darmflora haben, gestaltet sich der Nachweis schwierig: Es muss festgestellt werden, ob *E. faecalis* Symbioflor 1 unter den gefunden Stämmen ist.

Für den spezifischen Nachweis von Bakterien lässt sich die quantitative PCR nutzen. Indem für einen Stamm spezifische Primer gewählt werden, kann gezielt DNA dieses Keims amplifiziert und nachgewiesen werden.

Um für den Stamm spezifische Primer entwerfen zu können, benötigt man jedoch Kenntnisse über Sequenzen, die nur dieser Stamm besitzt. Diese sind nun verfügbar.

Mittlerweile ist auf diese Weise bei Versuchspersonen nachgewiesen worden, dass *E. faecalis* Symbioflor 1 die Magen-Darm-Passage überlebt und in der Tat zumindest einige Tage nach oraler Einnahme in der Darmflora persistiert und ausgeschieden wird (Domann, persönliche Kommunikation).

Weitere Setups sind denkbar. Zum Beispiel könnte man testen, inwieweit verschiedene Nahrungsmittel einen Einfluss auf die Dichte und Dauer der

Kolonisation mit *E. faecalis* Symbioflor 1 haben. Auch die Untersuchung, in welchen Darmabschnitten er sich vermehrt aufhält, ist so vorstellbar.

Auch wenn es nicht wahrscheinlich ist, könnte bei einer Enterokokken-Infektion auf diese Weise festgestellt werden, ob sie von diesem Stamm ausgelöst wird.

In vitro könnte getestet werden, ob der Stamm andere Enterokokken verdrängt.

5.6 Putative probiotische Eigenschaften

Mit *E. faecalis* V583 steht ein pathogenes *Enterococcus*-Isolat zur Verfügung (Paulsen et al. 2003), mit *E. faecalis* OG1RF ein Laborstamm (Bourgogne et al. 2008) und mit *E. faecalis* 62 ein kommensales Isolat eines Säuglings (D. A. Brede et al. 2011).

Mit dem Genom von *E. faecalis* Symbioflor 1 ist nun erstmalig ein Genom eines probiotisch verwendeten *E. faecalis* öffentlich verfügbar.

Dies bietet eine Reihe von Möglichkeiten für die probiotische Forschung.

Anfangs wurde erwähnt, dass bezüglich Probiotika gewisse Problematiken bestehen. Als Lebensmittel verkauft sollten sie der Förderung der "Wellness" dienen, ohne dass hierbei eine gesundheitlich förderliche Wirkung bewiesen werden musste (Schulze et al. 2008). Um solcherlei suggestiver Werbung entgegenzutreten, verabschiedeten europäisches Parlament und Rat eine Verordnung, die sogenannte "Health-Claims-Verordnung". Hiernach dürfen Lebensmittel nur dann als gesundheitsförderlich vermarktet werden, wenn ein solcher Nutzen wissenschaftlich bewiesen ist (Verordnung (EG) NR. 1924/2006)

Wenn Probiotika als Arzneimittel vermarktet werden sollen, gilt das Arzneimittelgesetz. Hiernach sind Arzneimittel Stoffe,

"[...]1. die zur Anwendung im oder am menschlichen oder tierischen Körper bestimmt sind und als Mittel mit Eigenschaften zur Heilung oder Linderung oder zur Verhütung menschlicher oder tierischer Krankheiten oder krankhafter Beschwerden bestimmt sind oder

2. die im oder am menschlichen oder tierischen Körper angewendet oder einem Menschen oder einem Tier verabreicht werden können, um entweder

- a) die physiologischen Funktionen durch eine pharmakologische, immunologische oder metabolische Wirkung wiederherzustellen, zu korrigieren oder zu beeinflussen oder
- b) eine medizinische Diagnose zu erstellen." (§2 AMG, Absatz 1, 2012)

Egal ob Probiotika also als Lebensmittel oder als Medikament verkauft werden, eine gesundheitsfördernde oder krankheitsverhindernde Wirkung ist wünschenswert.

Wie kann eine solche Wirkung aussehen?

In Bezug auf die probiotische Wirkung von Bakterien gibt es zahlreiche Vorstellungen und Modelle. Ein Mechanismus ist die Interaktion mit dem Darmepithel (Schulze, 2008).

Das annotierte Genom von *E. faecalis* Symbioflor 1 bietet die Basis, gezielt nach Proteinen zu suchen, die entweder sezerniert werden oder an der Oberfläche des Bakteriums exprimiert werden. Diese haben das Potential, mit den Darmepithelien und anderen Wirtszellen interagieren zu können und auf diese Weise eine Wirkung, zum Beispiel eine immunmodulatorische, zu entfalten.

Mit Augur wurden 4 potentiell sezernierte, für *E. faecalis* Symbioflor 1 spezifische Proteine identifiziert (André Billion et al. 2006).

Solche Proteine, insbesondere solche, die spezifisch für *E. faecalis* Symbioflor 1 als probiotischem Bakterium sind, könnten gezielt isoliert und in Zellkulturen angewendet werden, um deren Wirkung auf eukaryontische Zellen zu testen. Insbesondere könnte getestet werden, welche Wirkung diese Proteine auf die Durchlässigkeit von Epithelbarrieren haben.

In der Pathogenese von chronisch entzündlichen Darmerkrankungen spielen Lecks in den Tight Junctions des Darmepithels eine große Rolle (Teshima & Meddings 2008). Für verschiedene probiotisch verwendete Laktobazillen wurde gezeigt, dass sie die Durchlässigkeit des Darmepithels verhindern können (Sánchez et al. 2010; Schlee et al. 2008). Unter Umständen könnten auch Proteine dieses Stammes gefunden werden, die eine vergleichbare Wirkung haben. Unter Umständen könnte diese später sogar gezielt hergestellt und als Medikament verabreicht werden.

Um zu bestimmen, ob ein bestimmtes Proteinen überhaupt eine Wirkung hat, können, auf Basis der Genomsequenz, Deletionsmutanten hergestellt werden. Diese

können mittels Wachstumskurven und in Infektionsversuchen mit dem Wildtyp verglichen werden.

Wenn besser bekannt ist, welche Faktoren eine gesundheitsfördernde Wirkungen ausmachen, kann unter Umständen sogar versucht werden, Mutanten zu erzeugen, die die entsprechenden Faktoren vermehrt exprimieren, um die probiotische Wirkung zu verstärken.

Desweiteren können mit dem annotierten Genom als Referenz Transkriptions-Analysen durchgeführt werden. Auf diese Weise lässt sich untersuchen, welche Proteine der Keim unter verschiedenen Bedingungen exprimiert. So kann die Fragestellung untersucht werden, inwieweit der Wirt und dessen Mikrobiota Einfluss auf das Probiotikum hat.

Verschiedene Bakterien beeinflussen sich gegenseitig in ihrem Wachstum und ihrem Metabolismus, auch dies ist als probiotische Wirkungsweise diskutiert worden (Bisanz & Reid 2011).

Mit Transkriptions-Analysen ließe sich ebenso feststellen, welche Pathogenitätsfaktoren letztendlich überhaupt exprimiert werden.

In eine ähnliche Richtung wie Transkriptions-Analysen gehen metabolische Untersuchungen. Ein Teil der Wirkung probiotischer Bakterien wird darauf zurückgeführt, dass sie durch ihre Metabolite eine Millieu-Veränderung im Darm hervorrufen (Schulze, 2008).

In der komparativen Genom-Analyse fielen zwei Loci auf, an denen *E. faecalis* Symbioflor 1 zusätzliche Gene zum Transport und zum Metabolismus führte.

Der erste Locus betrifft ein Laktose / Cellobiose-spezifisches Zucker-Transportsystem. Dies suggeriert, dass der Keim sich ein Stück weit darauf eingestellt hat, vermehrt Laktose oder Cellobiose aufzunehmen.

Der zweite Locus ist möglicherweise durch eine Integrase in das Genom integriert worden. Er enthält unter anderem ein Zucker-Transport-System sowie eine Glukosidase / Galaktosidase. Zudem sind ein potentiell Zellwand-Protein und Restriktionsenzyme enthalten. Es ist denkbar, dass die Integration dieses Locus dem Keim einen lokalen Vorteil gegenüber anderen Enterokokken in Zucker-haltigen Lebensräumen verschafft.

Für beide Loci gilt einschränkend, dass solche Vermutungen nur mit Vorsicht gestellt werden können. Letztendlich wurden die Annotationen auf der Basis von Sequenz-Homologien getroffen. Da der Cutoff nicht bei 100% Identität gesetzt wurde, kann es natürlich sein, dass gerade aktive Zentren der betreffenden Proteine außer Kraft gesetzt oder in ihrer Spezifität verändert wurden. Zudem ist nicht sicher, ob die Proteine überhaupt exprimiert werden. In diesem Zusammenhang wäre wieder eine Transkriptionsanalyse hilfreich.

Zusammenfassend bietet die Genomsequenz von *E. faecalis* Symbioflor 1 eine Grundlage, von der aus weitere in die Tiefe gehende Forschungen gestartet werden können.

5.7 Ausblick

Mit dem fast vollständig sequenzierten Genom des *E. faecalis* Symbioflor 1 steht nun ein weiteres Genom dieser Bakterienspezies zur Verfügung.

Die Genomsequenz wurde als *.embl-Datei formatiert und in die Datenbank des European Molecular Biology Laboratory in Heidelberg submittiert (<http://www.ebi.ac.uk/embl/index.html>, submission ds: 85621).

Zukünftig bieten sich dadurch folgende Möglichkeiten:

1. Zukünftig sequenzierte *E. faecalis*-Stämme können vielleicht ein Stück leichter assembliert, geschlossen und annotiert werden, indem *E. faecalis* Symbioflor 1 als Referenz dient.

Mittlerweile wird Next-Generation-Sequenzierungs-Technologie immer kostengünstiger und damit weiträumiger verfügbar. Dadurch können in Zukunft auch neue *E. faecalis*-Isolate, sowohl kommensale als auch pathogene, einfacher sequenziert werden. Bisher vermag es jedoch keine Technologie, ein prokaryontisches Genom in einem Stück zu sequenzieren. Jede bisher verfügbare Technologie liefert nur eine Menge (im Vergleich zum gesamten Genom) relativ kurzer Reads. Daher sind in *de novo*-Assemblies in der Regel immer Gaps zu erwarten. Zur Ausrichtung der Contigs eines unbekannten Genoms und zum

Schließen der Gaps zwischen diesen sind Referenzgenome eine wertvolle Ressource.

Im Rahmen dieser Arbeit zeigte sich, dass die Gaps, die stark homolog zum Referenzgenom *E. faecalis* V583 waren, in aller Regel deutlich einfacher und schneller zu schließen waren. Insbesondere in repetitiven Bereichen kommt der Wert einer guten Referenz deutlich zum Tragen: Zum Beispiel konnten die rRNA-Kassetten durch Verwendung von Primern geschlossen werden, die zuvor für die rRNA-Kassetten anderer Bakterien entworfen wurden. Zuvor hatte die Sequenzierung dieser Bereiche aufgrund der repetitiven Natur durchaus Probleme bereitet.

So wie im Rahmen dieser Arbeit das Genom von *E. faecalis* V583 als Referenz zum ordnen der Contigs und zur Annotation der Gene von *E. faecalis* Symbioflor 1 diente, kann vielleicht auch das Genom von *E. faecalis* Symbioflor 1 als Referenz für einen zukünftig sequenzierten Stamm dienen. Die Erfahrung zeigt, dass dieser Prozess umso einfacher ist, je näher neuer Stamm und Referenzstamm miteinander verwandt sind. Naturgemäß sollte also jeder sequenzierte Stamm ein wenig dazu beitragen, dass zukünftige Stämme leichter zu sequenzieren und zu interpretieren sind.

2. Der Stamm *E. faecalis* Symbioflor 1 kann als Referenz zur Einordnung neu sequenzierter Enterokokken dienen. Wenn bei einem neuen Stamm unklar ist, inwieweit dieser pathogen ist, kann seine Sequenz mit der von bekannten Enterokokken verglichen werden, um einen Überblick zu erhalten.

Anders herum ist es nicht immer von vorn herein klar, was Pathogenität ausmacht. Um Gene zu finden, welche verdächtig für Pathogenität sind, bietet es sich zum Beispiel an, die Genome von vielen pathogenen Stämmen und vielen apathogenen Stämmen zu vergleichen. Über komparativ-statistische Analysen lassen sich so möglicherweise Gene finden, die vermehrt in pathogenen Stämmen enthalten sind und somit eventuell Pathogenität vermitteln.

Wenn diese Bereiche bekannt und gut charakterisiert sind, lassen sich eventuell sogar PCR-Primer entwickeln, mit denen ein unbekanntes Isolat dann schnell auf Pathogenität getestet werden kann.

6 Zusammenfassung

Enterococcus faecalis Symbioflor 1 (DSM 16431) ist ein Bakterienstamm, welcher Hauptbestandteil eines probiotischen Medikamentes mit dem Namen Symbioflor 1 ist.

Enterokokken gelten als fakultativ pathogene Gattung. Um die Bedrohung durch mögliche Pathogenitätsfaktoren auszuschließen und um den Stamm - als Inhaltsstoff eines Medikamentes - genauer zu definieren, wurde deshalb die Entscheidung getroffen, das Genom des *E. faecalis* Symbioflor 1 zu sequenzieren.

Zunächst wurde eine Plasmid-Shotgun-Bibliothek des Stamms sequenziert und assembliert. Hieraus resultierte ein Assembly mit 702 Contigs (Domann et al. 2007). Um die Aussagekraft bezüglich der Pathogenitätsfaktoren zu erhöhen und um den Stamm genauer zu untersuchen, sollte der Stamm möglichst komplett sequenziert werden und die Lücken zwischen den Contigs geschlossen werden.

Aufgrund der hohen Anzahl der Contigs wurde daher zunächst von der Firma Roche eine 454-Sequenzierung der chromosomalen DNA des Stammes durchgeführt. Mit dem Newbler-Assembler wurden 206.726 dieser 454-Reads zusammen mit 13.131 der Sanger-Reads der Plasmid-Bibliothek zu 69 Contigs > 500 bp assembliert. Von diesen konnten 45 am Referenzstamm *E. faecalis* V583 ausgerichtet werden.

Die Lücken zwischen diesen Contigs wurden sukzessive mittels PCRs geschlossen. Für große Lücken wurde zusätzlich eine Fosmid-Shotgun-Bibliothek der chromosomalen DNA des *E. faecalis* Symbioflor 1 mit 1.440 Klonen erzeugt, deren Inserts ansequenziert werden konnten um die Ausrichtung von Contigs zueinander festzustellen.

Das Resultat ergab ein zirkuläres Genom von 2.790.147 bp mit zwei zu diesem Zeitpunkt noch offenen Lücken.

Die Annotation des Genomes erfolgte halbautomatisch mit GenDB, wobei wiederum das Genom von *E. faecalis* V583 als Referenz diente.

Insgesamt besitzt der Stamm 2.714 codierende Sequenzen (CDS), welche im Schnitt 302 Aminosäuren codieren. Es sind 12 rRNA-Gene vorhanden, die sich auf 4 Loci verteilen, und 63 tRNA-Gene. Der MLST-Typ des Stammes ist 248.

Über phylogenetische Analysen konnte eine relativ nahe Verwandtschaft zu den Stämmen *E. faecalis* V583, *E. faecalis* OG1RF und *E. faecalis* 62 gezeigt werden. Es konnten 189 für *E. faecalis* *Symbioflor* 1 spezifische Gene gefunden werden.

Die Analyse bestätigte die weitgehende Abwesenheit von putativen Pathogenitätsfaktoren. Der Stamm kann daher als unbedenklich angenommen werden und wurde in Folge dessen von Risikogruppe 2 auf 1 herabgestuft.

Die nahezu komplette Genomsequenz bietet zukünftig die Möglichkeit, den Keim anhand spezifischer Sequenzen molekularbiologisch nachzuweisen sowie Transkriptions-Analysen durchzuführen und gezielt Mutanten zu erstellen, um die Wirkung von unbekannten Proteinen zu überprüfen, insbesondere im Hinblick darauf, möglicherweise probiotisch wirksame Faktoren zu entdecken.

Des Weiteren kann er als Referenz für zukünftig sequenzierte Enterokokken-Stämme dienen.

7 Summary

Enterococcus faecalis Symbioflor 1 (DSM 16431) is a bacterial strain which is the main component of a probiotic medicament named Symbioflor 1.

Enterococci are a facultative pathogenic genus. In order to exclude the danger caused by putative pathogenic traits and to further define the strain - as an ingredient of a medicament - it was decided to sequence the genome of *E. faecalis* Symbioflor 1.

Initially a plasmid-shotgun-library of the strain was sequenced and assembled, resulting in 702 contigs (Domann et al. 2007)

To increase the validity concerning the lack of pathogenic traits and to further investigate the strain, the strain should be sequenced as completely as possible and the gaps between contigs should be closed.

Because of the high number of contigs, the next step was a 454-sequencing of chromosomal DNA of the strain done by the Roche company. Using the Newbler assembler an assembly of 69 contigs > 500 bp was created by including 206,726 of said 454-reads and 13,131 of the Sanger-reads of the forementioned plasmid-library. Of those contigs, 45 could be mapped to the reference strain *E. faecalis* V583.

The gaps between those contigs were successively closed by PCR. For larger gaps a fosmid-shotgun-library of 1,440 clones was created. The inserts of said library were sequenced in order to align contigs to each other.

The result was a circular genome of 2,790,147 bp including two gaps that are still open

The genome sequence was then annotated semi-automatically using GenDB. *E. faecalis* V583 was used as a reference for annotation.

The strain possesses 2,714 coding sequences (CDS), each coding on average for 302 amino acids. Additionally, there are 12 rRNA-genes that are distributed among 4 loci, and 63 tRNA-genes. The MLST-type of the strain is 248.

The phylogenetic analysis revealed a close relationship to the strains *E. faecalis* V583, *E. faecalis* OG1RF and *E. faecalis* 62. 189 genes could be found that were specific for *E. faecalis* Symbioflor 1.

The analysis confirmed the absence of most putative pathogenic traits. Therefore the strain can be considered safe and it was downgraded from BSL (biosafety level) 2 to BSL1.

The almost complete genome sequence will provide the opportunity to detect the germ on the basis of specific sequences using molecular biological techniques. It will also enable transcription analyses and the creation of mutants in order to investigate the role of unknown proteins, especially to discover possible probiotic traits.

Additionally, the sequence can serve as a reference for future *Enterococcus* strains.

8 Eidesstattliche Erklärung

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.“

**Der Lebenslauf wurde aus der elektronischen
Version der Arbeit entfernt.**

**The curriculum vitae was removed from the
electronic version of the paper.**

10 Danksagung

Zuerst möchte ich mich bei meinem Doktorvater Prof. Dr. Eugen Domann bedanken, der mir dieses Thema zur Verfügung gestellt hat und mir immer zur Seite stand, wenn es Fragen oder Probleme gab. Außerdem danke ich für die stets schnellen und ergebigen Korrekturen meiner Arbeit.

Als nächstes möchte ich Prof. Chakraborty danken, der mir immer wieder, insbesondere bei der Arbeit mit den Fosmiden, mit Ideen und Hinweisen aushalf. Ebenso für Hilfe bei der Arbeit mit Fosmiden und Unterstützung zur Genomsequenzierung allgemein bin ich Dr. Torsten Hain zu Dank verpflichtet.

Für exzellente Betreuung im Labor, gute Tipps, den einen oder anderen Puffer und manchmal schier endlose Geduld bedanke ich mich herzlich bei Silke Zechel-Gran, die mir vor allem bei der Arbeit mit PCRs stets zur Seite stand und auch sonst immer hilfsbereit war. Auch Alexandra Amend möchte ich für ihre Geduld und Hilfe danken. Vor allem die Erstellung der Fosmidbank hätte ohne sie wahrscheinlich nicht geklappt. Ich danke auch Claudia Zörb sowie Nelly und Juri Schklarenko, die mir ebenfalls des Öfteren im Labor halfen, vor allem bei Sequenzierungen von PCRs aber auch bei Isolierungen von DNA. Isabell Trur, Kirsten Bommersheim und Manuela Pfeiffer danke ich vor allem für ihre Unterstützung bei Verwendung des PFGEs. Für immer schnelle Bestellung der Primer und für äußerst hilfreiches Know-how bei der Isolierung danke ich Martina Hudel.

Vor allem für moralische Unterstützung sowie die eine oder andere fachliche Diskussion möchte ich Dr. Svetlin Tchachalbachiev, Dr. Can Imirzalioglu, Martin Leustik und Maria Loose danken.

Desweiteren möchte ich mich bei Carsten Künnen und André Billion bedanken, die für mich die Bioinformatik ein wenig auf das Niveau eines "Mausschubers" runterbringen konnten und deren Programme für diese Arbeit unentbehrlich waren.

Zuletzt danke ich natürlich meiner lieben Ann-Kathrin für viel, viel Geduld und Energie, die sie für mich aufbrachte, und fürs Korrekturlesen meiner Arbeit. Und ich danke meiner Familie, die mir in dieser Zeit geduldig Rückhalt gab.

11 Literaturverzeichnis

- Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10.
- AMG, 2012. “Arzneimittelgesetz in der Fassung der Bekanntmachung vom 12. Dezember 2005 (BGBl. I S. 3394), das zuletzt durch Artikel 1 der Verordnung vom 16. Juli 2012 (BGBl. I S. 1534) geändert worden ist.”
- Angiuoli, S. V & Salzberg, S.L., 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics (Oxford, England)*, 27(3), pp.334–42.
- Apweiler, R. et al., 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic acids research*, 29(1), pp.37–40.
- Aziz, R.K. et al., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9, p.75.
- Badger, J.H. & Olsen, G. J., 1999. CRITICA: coding region identification tool invoking comparative analysis. *Molecular Biology and Evolution*, 16(4), pp.512–524.
- Bairoch, Amos et al., 2004. Swiss-Prot: juggling between evolution and stability. *Briefings in bioinformatics*, 5(1), pp.39–55.
- Barnes, W.M., 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences of the United States of America*, 91(6), pp.2216–20.
- Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), pp.1709–12.

- Billion, André et al., 2006. Augur--a computational pipeline for whole genome microbial surface protein prediction and classification. *Bioinformatics (Oxford, England)*, 22(22), pp.2819–20.
- Bisanz, J.E. & Reid, G., 2011. Unraveling how probiotic yogurt works. *Science translational medicine*, 3(106), p.106ps41.
- Bizzini, A. et al., 2010. Glycerol is metabolized in a complex and strain-dependent manner in *Enterococcus faecalis*. *Journal of bacteriology*, 192(3), pp.779–85.
- Bizzini, A. et al., 2009. The *Enterococcus faecalis* superoxide dismutase is essential for its tolerance to vancomycin and penicillin. *The Journal of antimicrobial chemotherapy*, 64(6), pp.1196–202.
- Bøhle, L.A. et al., 2010. Identification of proteins related to the stress response in *Enterococcus faecalis* V583 caused by bovine bile. *Proteom Science*.
- Bourgogne, A. et al., 2008. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome biology*, 9(7), p.R110.
- Brede, D.A. et al., 2011. Complete genome sequence of the commensal *Enterococcus faecalis* 62, isolated from a healthy Norwegian infant. *Journal of bacteriology*, 193(9), pp.2377–8.
- Brzuszkiewicz, E. et al., 2011. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Enter-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Archives of microbiology*, 193(12), pp.883–91.
- Camacho, C. et al., 2009. BLAST+: architecture and applications. *BMC bioinformatics*, 10, p.421.
- Chanos, P. & Williams, D.R., 2011. Anti-*Listeria* bacteriocin-producing bacteria from raw ewe's milk in northern Greece. *Journal of applied microbiology*.

- Chow, J.W. et al., 1993. Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. *Antimicrobial agents and chemotherapy*, 37(11), pp.2474–7.
- Christoffersen, T.E. et al., 2012. In vitro comparison of commensal , probiotic and pathogenic strains of *Enterococcus faecalis*. *British Journal of Nutrition*, pp.1–11.
- Copeland, A. et al., 2007. Complete sequence of chromosome of *Lactobacillus reuteri* DSM 20016.
- Courvalin, Patrice, 2006. Vancomycin resistance in gram-positive cocci. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 42 Suppl 1, pp.S25–34.
- Dahl, K.H. et al., 1999. Heterogeneity in the vanB gene cluster of genomically diverse clinical strains of vancomycin-resistant enterococci. *Antimicrobial agents and chemotherapy*, 43(5), pp.1105–10.
- Darling, A.C.E. et al., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7), pp.1394–403.
- Dear, S. & Staden, R., 1992. A standard file format for data from DNA sequencing instruments. *DNA sequence : the journal of DNA sequencing and mapping*, 3(2), pp.107–10.
- Delcher, A., 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, 27(23), pp.4636–4641.
- Domann, E. et al., 2007. Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic *Enterococcus faecalis* strain Symbioflor 1. *International journal of medical microbiology : IJMM*, 297(7-8), pp.533–9.
- Dunny, G.M., 1990. Genetic functions and cell-cell interactions in the pheromone-inducible plasmid transfer system of *Enterococcus faecalis*. *Molecular microbiology*, 4(5), pp.689–96.

- Eugen Domann, 1992. *Molekulargenetische Untersuchungen der Virulenzfaktoren Phospholipase (plcA), Listeriolysin (lisA), Protease (mpl) und Aktin-Bindung (actA) und deren koordinierte Regulation durch den positiven Regulator PrfA von Listeria monocytogenes*. Julius-Maximilians-Universität Würzburg.
- Europäisches Parlament und Rat, 2006. *Verordnung (EG) NR. 1924/2006 der europäischen Parlaments und des Rates vom 20. Dezember 2006 über nährwert- und gesundheitsbezogene Angaben über Lebensmittel* ◀ (ABl. L 404 vom 30.12.2006, S. 9),
- Felsenstein, J., 1985. Confidence Limits On Phylogenies: An Approach Using The Bootstrap. *Evolution*, 39, pp.783–791.
- Foulquié Moreno, M.R. et al., 2006. The role and application of enterococci in food and health. *International journal of food microbiology*, 106(1), pp.1–24.
- Frankenberg, L., Brugna, M. & Hederstedt, L., 2002. Enterococcus faecalis Heme-Dependent Catalase. , 184(22), pp.6351–6356.
- Ghai, R., Hain, Torsten & Chakraborty, Trinad, 2004. GenomeViz: visualizing microbial genomes. *BMC bioinformatics*, 5, p.198.
- Goldin, B.R. & Gorbach, S.L., 2008. Clinical indications for probiotics: an overview. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 46 Suppl 2(Supplement 2), pp.S96–100; discussion S144–51.
- Grissa, I., Vergnaud, G. & Pourcel, C., 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*, 35(Web Server issue), pp.W52–7.
- Habermann, W. et al., 2002. [Reduction of acute recurrence in patients with chronic recurrent hypertrophic sinusitis by treatment with a bacterial immunostimulant (Enterococcus faecalis Bacteriae of human origin]. *Arzneimittel-Forschung*, 52(8), pp.622–7.

- Habermann, W. et al., 2001. [The effect of a bacterial immunostimulant (human *Enterococcus faecalis* bacteria) on the occurrence of relapse in patients with]. *Arzneimittel-Forschung*, 51(11), pp.931–7.
- Hancock, L.E. & Gilmore, M.S., 2002. The capsular polysaccharide of *Enterococcus faecalis* and its relationship to other polysaccharides in the cell wall. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), pp.1574–9.
- Hendrickx, A.P. a et al., 2009. LPxTG surface proteins of enterococci. *Trends in microbiology*, 17(9), pp.423–30.
- Hillis, D.M. & Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of biology*, 66(4), pp.411–53.
- Ho, C.-C. et al., 2011. Rapid identification and validation of specific molecular targets for detection of *Escherichia coli* O104:H4 outbreak strain by use of high-throughput sequencing data from nine genomes. *Journal of clinical microbiology*, 49(10), pp.3714–6.
- Hof, H. et al., 2005. *Medizinische Mikrobiologie*, Duale Reihe.
- Höffken, G. et al., 2009. Epidemiologie, Diagnostik, antimikrobielle Therapie und Management von erwachsenen Patienten mit ambulant erworbenen unteren Atemwegsinfektionen sowie ambulant erworbener Pneumonie – Update 2009. *Pneumologie (Stuttgart, Germany)*, 63(10), pp.e1–68.
- Imirzalioglu, C. et al., 2008. Hidden pathogens uncovered: metagenomic analysis of urinary tract infections. *Andrologia*, 40(2), pp.66–71.
- Jett, B.D., Huycke, M.M. & Gilmore, M.S., 1994. Virulence of enterococci. *Clinical microbiology reviews*, 7(4), pp.462–78.
- Jurkovic, D. et al., 2006. Identification and characterization of enterococci from bryndza cheese. *Letters in applied microbiology*, 42(6), pp.553–9.

- Kim, Y.J., Kim, E.-H. & Hahm, K.B., 2012. Oxidative stress in inflammation-based gastrointestinal tract diseases: challenges and opportunities. *Journal of gastroenterology and hepatology*, 27(6), pp.1004–10.
- Kreft, B. et al., 1992. Aggregation substance of *Enterococcus faecalis* mediates adhesion to cultured renal tubular cells. *Infection and immunity*, 60(1), pp.25–30.
- Kuenne, C.T. et al., 2007. GECO--linear visualization for comparative genomics. *Bioinformatics (Oxford, England)*, 23(1), pp.125–6.
- Ley, R.E. et al., 2009. Worlds within worlds : evolution of the vertebrate gut microbiota. , 6(10), pp.776–788.
- Löffler, Georg; Petrides, P.E., 2002. *Biochemie und Pathobiochemie*, Springer.
- Lowe, T.M. & Eddy, S R, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research*, 25(5), pp.955–64.
- Makarova, K.S. et al., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct*, 1, p.7.
- Martínez-Marcos, F.J. et al., 2009. Endocarditis por enterococo: análisis multicéntrico de 76 casos. *Enfermedades Infecciosas y Microbiología Clínica*, 27(10), pp.571–579.
- Mellmann, A. et al., 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PloS one*, 6(7), p.e22751.
- Meyer, F., 2003. GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Research*, 31(8), pp.2187–2195.
- Murray, B.E., 1990. The life and times of the *Enterococcus*. *Clinical microbiology reviews*, 3(1), pp.46–65.

- Nallapareddy, S.R. et al., 2011. Relative contributions of Ebp Pili and the collagen adhesin ace to host extracellular matrix protein adherence and experimental urinary tract infection by *Enterococcus faecalis* OG1RF. *Infection and immunity*, 79(7), pp.2901–10.
- Nieto-Arribas, P. et al., 2011. Enterococcus populations in artisanal Manchego cheese: Biodiversity, technological and safety aspects. *Food Microbiology*, 28(5), pp.891–899.
- Pajon, A. et al., 2010. The genome sequence of *Enterococcus* sp. 7L76. *unpublished*.
- Palmer, K.L. & Gilmore, M.S., 2010. Multidrug-resistant enterococci lack CRISPR-cas. *mBio*, 1(4).
- Paulsen, I.T. et al., 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science (New York, N.Y.)*, 299(5615), pp.2071–4.
- Pillar, C.M. & Gilmore, M.S., 2004. Enterococcal Virulence - Pathogenicity Island of *E. faecalis*. , pp.2335–2346.
- Pitcher, D.G., Saunders, N.A. & Owen, R.J., 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology*, 8(4), pp.151–156.
- Poh, C.H., Oh, H.M.L. & Tan, A.L., 2006. Epidemiology and clinical outcome of enterococcal bacteraemia in an acute care hospital. *The Journal of infection*, 52(5), pp.383–6.
- Pritchard, G.G. & Wimpenny, J.W., 1978. Cytochrome formation, oxygen-induced proton extrusion and respiratory activity in *Streptococcus faecalis* var. zymogenes grown in the presence of haematin. *Journal of general microbiology*, 104(1), pp.15–22.
- Punta, M. et al., 2012. The Pfam protein families database. *Nucleic acids research*, 40(Database issue), pp.D290–301.

- Quintiliani, R. & Courvalin, P., 1994. Conjugal transfer of the vancomycin resistance determinant vanB between enterococci involves the movement of large genetic elements from chromosome to chromosome. *FEMS microbiology letters*, 119(3), pp.359–63.
- Rich, R.L. et al., 1999. Ace is a collagen-binding MSCRAMM from *Enterococcus faecalis*. *The Journal of biological chemistry*, 274(38), pp.26939–45.
- Ritchey, T. W., H. W. Seeley, J., 1974. Cytochromes in *Streptococcus faecalis* var. zymogenes Grown in a Haematin-containing Medium. *Journal of General Microbiology*, 85, pp.220–228.
- Robert-Koch-Institut, 2011. *Abschließende Darstellung und Bewertung der epidemiologischen Erkenntnisse im EHEC O104:H4 Ausbruch, Deutschland 2011*, Berlin.
- Ronaghi, M., 2001. Pyrosequencing Sheds Light on DNA Sequencing
Pyrosequencing Sheds Light on DNA Sequencing. , pp.3–11.
- Rosenkranz, W. & Grundmann, E., 1994. [Immunomodulator action of living, nonpathogenic *Enterococcus faecalis* bacteria from humans]. *Arzneimittel-Forschung*, 44(5), pp.691–5.
- Ruiz-Garbajosa, P. et al., 2006. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *Journal of clinical microbiology*, 44(6), pp.2220–8.
- Sahm, D.F. et al., 1989. In vitro susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. *Antimicrobial Agents and Chemotherapy*, 33(9), pp.1588–1591.
- Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4), pp.406–25.

- Sánchez, B., Urdaci, M.C. & Margolles, A., 2010. Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa-bacteria interactions. *Microbiology (Reading, England)*, 156(Pt 11), pp.3232–42.
- Schaberg, D.R., Culver, D.H. & Gaynes, R.P., 1991. Major trends in the microbial etiology of nosocomial infection. *The American journal of medicine*, 91(3B), p.72S–75S.
- Schlee, M. et al., 2008. Probiotic lactobacilli and VSL#3 induce enterocyte beta-defensin 2. *Clinical and experimental immunology*, 151(3), pp.528–35.
- Schmiemann, G. et al., 2010. Epidemiologie , Diagnostik , Therapie und Management unkomplizierter bakterieller ambulant erworbener Harnwegsinfektionen bei erwachsenen Patienten. , (043).
- Schulze, J. et al., 2008. *Probiotika. Mikroökologie, Mikrobiologie, Qualität, Sicherheit und gesundheitliche Effekte* Sonderausg., Thieme-Verlag.
- Semedo, T. et al., 2003. Virulence Factors in Food , Clinical and Reference Enterococci : A Common Trait in the Genus ? , 22, pp.13–22.
- Shankar, N., Baghdayan, A.S. & Gilmore, M.S., 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. *Nature*, 417(6890), pp.746–50.
- Sonnenborn, U. & Greinwald, R., 1991. *Beziehungen zwischen Wirtsorganismus und Darmflora*, Stuttgart, New York: Schattauer Verlag.
- Steinweg, C., 2005. *Genomsequenzierung von L. Welshimeri und L. Seeligeri, zwei apathogenen Spezies des Genus Listeria*, VVB Lauferweiler.
- Tamura, K. et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), pp.2731–9.
- Tamura, K., Nei, Masatoshi & Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30), pp.11030–5.

- Tanabe, M. & Kanehisa, M., 2012. Using the KEGG Database Resource. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*, Chapter 1, p.Unit1.12.
- Tatusov, R.L., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1), pp.33–36.
- Tendolkar, P.M. et al., 2004. Enterococcal surface protein, Esp, enhances biofilm formation by *Enterococcus faecalis*. *Infection and immunity*, 72(10), pp.6032–9.
- Teng, F. et al., 2009. Further characterization of the epa gene cluster and Epa polysaccharides of *Enterococcus faecalis*. *Infection and immunity*, 77(9), pp.3759–67.
- Teshima, C.W. & Meddings, J.B., 2008. The Measurement and Clinical Significance of Intestinal Permeability.
- Thurlow, L.R., Thomas, V.C. & Hancock, L.E., 2009. Capsular polysaccharide production in *Enterococcus faecalis* and contribution of CpsF to capsule serospecificity. *Journal of bacteriology*, 191(20), pp.6203–10.
- Toledo-Arana, A. et al., 2001. The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. *Applied and environmental microbiology*, 67(10), pp.4538–45.
- Vebø, H.C. et al., 2010. Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. *PloS one*, 5(8), p.e12489.
- Voelkerding, K. V, Dames, S. a & Durtschi, J.D., 2009. Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*, 55(4), pp.641–58.
- Wardlaw, T. et al., 2010. Diarrhoea: why children are still dying and what can be done. *The Lancet*, 375(9718), pp.870–872.
- WHO, 2008. *The global burden of disease: 2004 update*, WHO Library Cataloguing-in-Publication Data.

- Wilkinson, L., 2012. Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE transactions on visualization and computer graphics*, 18(2), pp.321–31.
- Xu, Y., Murray, B.E. & Weinstock, G M, 1998. A cluster of genes involved in polysaccharide biosynthesis from *Enterococcus faecalis* OG1RF. *Infection and immunity*, 66(9), pp.4313–23.
- Zhou, Y. et al., 2011. PHAST: a fast phage search tool. *Nucleic acids research*, 39(Web Server issue), pp.W347–52.

12 Publikationen

Zum Stand der Abgabe der schriftlichen Dissertation waren noch 2 Lücken im Genom des *E. faecalis* Symbioflor 1 offen. Im Zeitraum zwischen Abgabe und Disputation gelang es, diese mittels eines zu diesem Zeitpunkt im Institut neu eingeführten Next Generation Sequencers, dem Illumina MiSeq, zu schließen. Die so gewonnene komplette Genomsequenz wurde in der ENA-Datenbank (European Nucleotide Archive, <http://www.ebi.ac.uk/ena>) unter der Accession-Nummer HF558530 hinterlegt und ist dort frei für die Öffentlichkeit zugänglich.

Die Sequenzierung des Genoms wurde mit einem kurzen Artikel in der Zeitschrift Genome Announcements (American Society of Microbiology) bekanntgegeben, der am 7. Februar 2013 veröffentlicht wurde. Siehe hierzu: Fritzenwanker, M. et al., 2013. Complete Genome Sequence of the Probiotic *Enterococcus faecalis* Symbioflor 1 Clone DSM 16431. Genome announcements, 1(1).

13 Anhang

13.1 Primer der ersten Runde Lückenschluss-PCRs (minimap)

EF1_phase1_gap1_L	TCAATGCCTAACGATGAACG	EF1_phase1_gap33_L	TCCTCTTGCTACCGGAAATC
EF1_phase1_gap1_R	ATTTTCATATGCTAAGCGGCG	EF1_phase1_gap33_R	TGAGAATGCTCCACGTCTTG
EF1_phase1_gap2_L	AACAAACGGTCATCGGTTTC	EF1_phase1_gap34_L	TATCAGCAACAATTGGCAGC
EF1_phase1_gap2_R	TTGAAATGCTTCCGACATTG	EF1_phase1_gap34_R	CAACATTTGAAAAATGGGGC
EF1_phase1_gap3_L	AAAATGAACTGGCAAGCGTC	EF1_phase1_gap35_L	AAAATTCAACGCCGTTCTTG
EF1_phase1_gap3_R	AAAGCTCAGGCCTAGGAAGC	EF1_phase1_gap35_R	ATATCAATGACTGGGAGCCG
EF1_phase1_gap4_L	ACGATTGATCCGAAACAAGG	EF1_phase1_gap36_L	TCCATTACGTTAGGGGTTGC
EF1_phase1_gap4_R	ATACCGCAAAAATGCCTCTG	EF1_phase1_gap36_R	GACCCCTTCTCAACGAATTG
EF1_phase1_gap5_L	ACGGGAACAAGGAAATGATG	EF1_phase1_gap37_L	TGGGCTAAAAACCAGTCCAC
EF1_phase1_gap5_R	TGGAAGGCTGTGGTTTTACC	EF1_phase1_gap37_R	TAGGAACAGCTGCAATGTCTG
EF1_phase1_gap6_L	AAAAGCATGTTCTGATGGGG	EF1_phase1_gap38_L	TACGCTAGAGCAATGGCTTG
EF1_phase1_gap6_R	CTAGACCGTTTCTCGATGCC	EF1_phase1_gap38_R	CCCATTATTCCGGCACTATC
EF1_phase1_gap7_L	CCTTCGACGGTGAAAAGAAC	EF1_phase1_gap39_L	TCATTTTAGCCGGTGGTAGC
EF1_phase1_gap7_R	TCTGCCCCATTAGCAAAAAC	EF1_phase1_gap39_R	ACACCGCCAGTCAGAGAATC
EF1_phase1_gap8_L	GAGGATGGAGTCCCTTTTGAG	EF1_phase1_gap40_L	ACCAGAGATGGCTCAACCAC
EF1_phase1_gap8_R	ATACCGTTCTCGTGGGTTTG	EF1_phase1_gap40_R	GTGGAAGAAATTCGAGCCAG
EF1_phase1_gap9_L	TTTAGAAACGGGTCCAACG	EF1_phase1_gap41_L	AGGGCTATCCGCTCTTCTTC
EF1_phase1_gap9_R	AATTATGGCGCTCTCCATTG	EF1_phase1_gap41_R	ATTGATGAAAACTTGCGGC
EF1_phase1_gap10_L	ACGAAGTACGCCGATGTTTC	EF1_phase1_gap42_L	TGAAGCAATGGAATTTGCTG
EF1_phase1_gap10_R	ATGGCTTCTGTAATGGCTGG	EF1_phase1_gap42_R	AGCAGGCAAAATGATGGAAC
EF1_phase1_gap11_L	GATGTTGCTCAATTGCATGG	EF1_phase1_gap43_L	CCCATTTCACCGAAGAAAAC
EF1_phase1_gap11_R	CCATAGAAGCTCGCTTGACC	EF1_phase1_gap43_R	CCGCCGAAGTTGAGTGATC
EF1_phase1_gap12_L	TTGCTACGCAAGATGAATGG	EF1_phase1_gap44_L	TAAGTGAGGTGTCCTTGCCC
EF1_phase1_gap12_R	TTGAAATTGCTGCTCGTCAC	EF1_phase1_gap44_R	ATACCAATGGTTGGTCTGGC
EF1_phase1_gap13_L	ACTTATGGCGGTTATGCTCG	EF1_phase1_gap45_L	TATGTGCTTTTTCGCTGTCTG
EF1_phase1_gap13_R	TTGCCACAAAAACACTCACC	EF1_phase1_gap45_R	TGAATCTGGCTATTGGCCTC
EF1_phase1_gap14_L	CAAATTCGCCCAATCAAAAC	EF1_phase1_gap46_L	ACCGATTTTGACTTCATGC
EF1_phase1_gap14_R	CTTTGTGAGCCAAGCCTTTC	EF1_phase1_gap46_R	TGCGTTATCTAAAGGGGTCTG
EF1_phase1_gap15_L	ACCAGCAATAATCGCAAACC	EF1_phase1_gap47_L	AGCGCTACACCTCTCTTTGC
EF1_phase1_gap15_R	CAGTTCAACTTGCGATGCTC	EF1_phase1_gap47_R	ATGACTGTGGAAGGCTCAGG
EF1_phase1_gap16_L	GCAGGTGTTGTGTATCACGC	EF1_phase1_gap48_L	CGCATCAAGTTCTGTGGTATG
EF1_phase1_gap16_R	GGCATGAATGTGAAACTCCC	EF1_phase1_gap48_R	AGCAAAGGTCAATCAATGGG
EF1_phase1_gap17_L	AGAAGTTGCGCATGTTACCC	EF1_phase1_gap49_L	TGCAAAAAGCCTGTTCTCTG
EF1_phase1_gap17_R	ATCATTACGGCGACTGATCC	EF1_phase1_gap49_R	AGAAGAAGTTGGCGCTTCAG
EF1_phase1_gap18_L	ATGGGGCGTTACAAAACTG	EF1_phase1_gap50_L	TTGCATATCCCCTAATTCCG
EF1_phase1_gap18_R	TGCATGTTATTGCCTTTTGC	EF1_phase1_gap50_R	CACGATTGATCGTTTGTACG
EF1_phase1_gap19_L	CGTTTTGGAAGTGGAGAAGG	EF1_phase1_gap51_L	ATTAACCACAGAAGGTGGCG
EF1_phase1_gap19_R	TGAACTAGGTCCCGAAGGTG	EF1_phase1_gap51_R	TAGTATTAGGTGGCGCAGGC
EF1_phase1_gap20_L	TGCAGAAAGCGTATTGCATC	EF1_phase1_gap52_L	CGACCAATGCCTGATTTAGC
EF1_phase1_gap20_R	TGAAAGGGCGGTGTCTTAAC	EF1_phase1_gap52_R	CTTTGATTCCCCCGTTTATG
EF1_phase1_gap21_L	AAGAACAACCAGGCGTTTTG	EF1_phase1_gap53_L	TTATTACATGTCTGCAAGGCG

EF1_phase1_gap21_R	AACGGTCTCTTGCTCACAAGC	EF1_phase1_gap53_R	GCGCTAGAAGAAATTGCAGC
EF1_phase1_gap22_L	AAATGCGAAAATCACCTTGTC	EF1_phase1_gap54_L	CTTTGTGAGCCAAGCCTTTC
EF1_phase1_gap22_R	GCTCGTAAAATTCGCTCGTC	EF1_phase1_gap54_R	ATGCTTCTGGCAAAATGACC
EF1_phase1_gap23_L	TTTAGAGGCGGCAAAAATTG	EF1_phase1_gap55_L	TCCAGGATTTTCACAGACCC
EF1_phase1_gap23_R	ATGGCTGCAAAAGAAGCAAG	EF1_phase1_gap55_R	AAGCAATTCCCGTTTTAGGC
EF1_phase1_gap24_L	ATAAAAACCACTGCGGCAAC	EF1_phase1_gap56_L	AGACACCCACTGGCGATAAC
EF1_phase1_gap24_R	TTCGGAAATACTCCCACCAG	EF1_phase1_gap56_R	CGCTTAACCTTGCGAATCC
EF1_phase1_gap25_L	ACAACCACTGGAAAGTTGGC	EF1_phase1_gap57_L	AGGTGTTTGGCTTGGTTTTG
EF1_phase1_gap25_R	GCAAAATGACGGCTCTAATCC	EF1_phase1_gap57_R	GGATTACACAAGGAAAGCG
EF1_phase1_gap26_L	TACAGCGAAATTACCCGACC	EF1_phase1_gap58_L	AAGCCCATAATCGTTGTTGC
EF1_phase1_gap26_R	ATGGGTAAGACCCCTTCGAC	EF1_phase1_gap58_R	AATTGAAACAGTTGGTGCCG
EF1_phase1_gap27_L	TCTCTGCTGTTCAAGCATCG	EF1_phase1_gap59_L	TCATTGTGCCGATTCTATC
EF1_phase1_gap27_R	GTTGCATCAACTGTATCGGC	EF1_phase1_gap59_R	TAGCGTAATAGTCGGGCACC
EF1_phase1_gap28_L	GATTGTTACCGTCGTGATG	EF1_phase1_gap60_L	TCAACGTGATTGCTTCTGC
EF1_phase1_gap28_R	ATTGCATTTGCTGGGATTTTC	EF1_phase1_gap60_R	AGAAAGCAACGATCACCCAC
EF1_phase1_gap29_L	TATGGGGATTTCTACGCAGC	EF1_phase1_gap61_L	AACGTGCCCCGCTTCATTGAC
EF1_phase1_gap29_R	TCATTCCGAAGAAATCGGTC	EF1_phase1_gap61_R	TTCCTGCTCCAACAGGATTC
EF1_phase1_gap30_L	CGTGTCATGCCAAGTTTTG	EF1_phase1_gap62_L	TTGTCGCTTTTGCTTTGATG
EF1_phase1_gap30_R	TTGGTTTATCGGGAGCATTC	EF1_phase1_gap62_R	AATCACCGTCTTTTGCTTCG
EF1_phase1_gap31_L	ATACTCACCCAATTGCGCTC	EF1_phase1_gap63_L	ACAGTGATTTTGCCATTCC
EF1_phase1_gap31_R	GGTTGGGACAACAAACCAAG	EF1_phase1_gap63_R	AACAAAGGCGTCATTATCCG
EF1_phase1_gap32_L	TTTCAAAGGTGGTAAAGCCG		
EF1_phase1_gap32_R	TAGCCCAAACGAATACCAGC		

13.2 Primer der zweiten Phase Lückenschluss-PCRs (minimap, Seqman, Primer3+)

EF1_phase2_gap1_pair1_L	ACGATTGATCCGAAACAAGG	EF1_ph2.4_gap25_L_sm	GGTCCGCTCCAACATATCA
EF1_phase2_gap1_pair2_R	TTCAGTAACCGCAAGGAAGG	EF1_ph2.4_gap25_R_sm	CAGTAGGGAATCTTCGGCAATGGA
EF1_phase2_gap2_pair5_L	GAAGTTCCCTAATCCATCAAGC	EF1_ph2.4_gap5_L_sm	ATCTTCGGCAGGAGCTAACAGTGA
EF1_phase2_gap2_pair1_R	ATACCGCAAAAATGCCTCTG	EF1_ph2.4_gap5_R_sm	TGACAAATGGGTAAAGCAATCTAA
EF1_phase2_gap3_pair3_L	TTTCACTGACCGCTTTGAAC	EF1_ph2.4_gap12_L_sm	TCGCTGCAAAATAAGAATGA
EF1_phase2_gap3_pair2_R	TGAGCTAAAGGTCCAGGCTC	EF1_ph2.4_gap12_R_sm	CACCTGTTGCTCCTACCAC
EF1_phase2_gap4_pair1_L	TAGGCCTTTATGGGCATCAG	EF1_ph2.4_gap15_L_sm	AAGAGAAAAGGGGTATGAAAATGG
EF1_phase2_gap4_pair4_R	ATAGCTCGTCTGCCTTCTGG	EF1_ph2.4_gap15_R_sm	TAACAAAACACTACTCTTCTATCA
EF1_phase2_gap5_pair1_L	GAGGATGGAGTCCCTTTTGAG	EF1_ph2.4_gap11_R_p+	CAGAACCCCCACATTCTTCT
EF1_phase2_gap5_pair3_R	TCCTACGGCTCCTAACACAC	EF1_ph2.4_gap19_L_p+	GGTACCATCTTGCGCTTCTT
EF1_phase2_gap6_pair1_L	CCCAACGATAGACAGTGGTG	EF1_ph2.4_gap6_L_sm	ATACTGGTGCCAATAATCAAGAGC
EF1_phase2_gap6_pair1_R	ACCAATCGTAGCTGCTTTTCG	EF1_ph2.4_gap6_R_sm	GGAAAGCCACGTGTAGAAATCAT
EF1_phase2_gap7_pair2_L	TCTCAATTTAGGAATGAGATGGTG	EF1_ph2.4_gap28a_L_sm	TTTAATCGTTGAAGCATCTACTCG
EF1_phase2_gap7_pair3_R	TGAAAAAGTGGATTTTGTAG	EF1_ph2.4_gap28b_R_sm	GCGAACGGTTATTAATAGTGAAGC
EF1_phase2_gap8_pair1_L	ATATGGTGCAAGTCAGCACG	EF1_ph2.4_gap26_L_sm	ATTTTGTGCATATTTGGCTAAGGT
EF1_phase2_gap8_pair1_R	ATACCGTTCTCGTGGGTTTG	EF1_ph2.4_gap26_R_sm	ATTTAATCATTGCAAGCCCACCAT
EF1_phase2_gap9_pair1_L	TTGCTACGCAAGATGAATGG	EF1_2.5_gp7.1_L_sm	GGATTTAAAAAGATACCGATAGTG
EF1_phase2_gap9_pair2_R	AGCGTTTCTGCAGGTTTAGC	EF1_2.5_gp7.2_L_sm	ATTGCGTCATCGTTACATCACTA
EF1_phase2_gap10_pair1_L	AGCCACAACGATGAAAAAAC	EF1_2.6_gp11_L_sm	TGCAAGAAATAAAGGCGGTAATGG
EF1_phase2_gap10_pair1_R	CTTTGTGAGCCAAGCCTTTC	EF1_2.6_gp8_L_sm	TTTCCACAATCACTCACTTCTTT

EF1_phase2_gap11_pair2_R	ACGGAACCAACCGAGTAATG	EF1_2.6_gp8_R_sm	AATTGAAAAAGGTAAGGAAAGAAA
EF1_phase2_gap11_pair5_L	AAGAGAGCAACGTACAATAAAG	EF1_2.6_gp12_L_sm	TAGTCCTTTGATGGGTAATAGTTT
EF1_phase2_gap12_pair3_R	AGCGATCGACGAAAAGATTG	EF1_2.6_gp23_L_sm	GACCACGTTCTGCCAAATCAAATG
EF1_phase2_gap12_pair6_L	TTGTAGGTAAGACACAAAAGTG	EF1_2.6_gp23_R_sm	TGATGAGAAAAGATTACAAGTCG
EF1_phase2_gap13_pair2_R	TTTTTCATTCCGCATCACTCC	EF1_2.6_gp11AB_L_sm	AAGTATTTGAATTGGTCTGTGC
EF1_phase2_gap13_pair3_L	CGTACGCAGTATGCAGGAAG	EF1_2.6_gp11AB_R_sm	TTACAGGAACCCCAGCAAATA
EF1_phase2_gap14_pair1_L	TCGATTTTACCGAGCGAAAG	EF1_2.6_gp24_L_h	TCCTGATTTCAATTCTGAGTTTGG
EF1_phase2_gap14_pair2_R	TAAGCGAGGGCTGTTCTTTC	EF1_2.6_gp26_L_sm	CAGGTGGGAATTGTATGTTGA
EF1_phase2_gap15_pair1_R	AACTTGGCCCCAAATAAACC	EF1_2.6_gp26_R_sm	AAATAAATAGAAAGGAGTGAGTT
EF1_phase2_gap15_pair3_L	GAGAAAAGGGGTATGAAAATGG	EF1_2.6_gp6_L_sm	TACGTGCCAAATCCAAATCTGTGA
EF1_phase2_gap16_pair1_L	TTTTTACTCGCCGTCTTTGG	EF1_2.7_gp17_L_sm	CATTGGTGCTTCTATTTA
EF1_phase2_gap16_pair3_R	CATCCCATCCAATATCACCC	EF1_2.7_gp17_R_sm	TCGGCTTATTCATCATTGGACAC
EF1_phase2_gap17_pair1_L	CTACCGGCCAATCGTAACTG	EF1_2.7_gp20_L_sm	CCAGCCAAACGGAGTCATTCAGA
EF1_phase2_gap17_pair1_R	ATCCATTGCCACGACTTCTC	EF1_2.7_gp20_R_sm	CCGGCACTATCTACCACATTTCT
EF1_phase2_gap18_pair1_L	ATATCAATGACTGGGAGCCG	EF1_2.7_uc102_A_sm	TTTGCGTATTCGTTTGATTTTCTT
EF1_phase2_gap18_pair1_R	GACCCCTTCTCAACGAATTG	EF1_2.7_uc102_B_sm	ATAAATGGCTCAGCAACTACAATG
EF1_phase2_gap19_pair1_L	ATTGGCTGATCTAACACGGC	EF1_2.7_uc091_A_sm	ACCGTCGTTTCAGGCTATCACT
EF1_phase2_gap19_pair3_R	TAGATCCGACTATTGCCTGC	EF1_2.7_uc091_B_sm	CCTTATATCGGCCGGCTTTTTGA
EF1_phase2_gap20_pair1_R	CCCATTATTCCGGCACTATC	EF1_2.7_uc084_A_sm	AACACTTATTTGCGCTCTATTTCA
EF1_phase2_gap20_pair5_L	GGCACTTTATCTACTGAAAATCAAATC	EF1_2.7_uc084_B_sm	TTGCGACGATTGGCTTAGAAACAC
EF1_phase2_gap21_pair2_L	ATGATTTGCAACGTCCAACC	EF1_2.7_gp18_L_sm	TGTCAGGACCTACTATACTTCTCT
EF1_phase2_gap21_pair2_R	ATTCGCTCGTGATGTTTTTC	EF1_2.7_gp18_R_sm	TTGCAGACATGAAAACCTCC
EF1_phase2_gap22_pair1_L	ATTCAATAGGCGAACCAACG	EF1_2.8_uc094_B_sm	AGTGCTGAAATGACGGCTAAATGG
EF1_phase2_gap22_pair1_R	GACGTTTGAAGAGAGCCCTG	EF1_2.8_gp17_L_sm	TTTTGTGCATATTTGGCTGAGGTA
EF1_phase2_gap23_pair1_R	TGCGTTATCTAAAGGGGTCG	EF1_2.8_gp17_R_sm	AGGGGAATAAATGAACGAGAACAG
EF1_phase2_gap23_pair3_L	ATCTTTTCGTCGATCGCTTC	EF1_2.8_uc084_B_sm	TTTGTTAGCAGGATGTAGCACGAC
EF1_phase2_gap24_pair1_R	ATACTTCCCCAAGTTGCACG	EF1_2.8_uc113_B_sm	ATCTTCGGCAGGAGCTAACAGTGA
EF1_phase2_gap24_pair2_L	TGATACGATCGTTGGCATTC	EF1_2.8_gp20_L_sm	ACAAACTGGAATCCTAGAAGCATC
EF1_phase2_gap25_pair1_R	TTATTACATGTCGCAAGGCG	EF1_2.8_uc102_B_sm	TAGAGGCAATGTTCCAGTAGTGA
EF1_phase2_gap25_pair2_L	CGTGCTTTTCTGACCATACG	EF1_2.8_uc121_A_sm	CAAAGTTGTATCCCCAGTAAATCA
EF1_phase2_gap26_pair1_L	CGTTACATCTGGTTGTGCG	EF1_2.8_gp16_L_sm	AACCATTTGTCAACGCTTCCTCA
EF1_phase2_gap26_pair2_R	AGTAGAAATCGCACGCCAAC	EF1_2.8_gp16_R_sm	TGTGATTGAAAAAGGTGGAGAAGA
EF1_phase2_gap27_pair1_L	TCCAGGATTTTACAGACCC	EF1_2.9_gp17_L_sm	TGTCTAAGTCCGTGCATTGTGAT
EF1_phase2_gap27_pair1_R	GTTCGGAAGGTAGCAGTTGG	EF1_2.9_gp17_R_sm	TTCTGAAACAATGGATACAACAA
EF1_phase2_gap28_pair1_R	GGAATTTGCGATTCTTTTCG	EF1_2.9_gp6_L_sm	GGTTAGTTTTTCTTGCCAGTTATT
EF1_phase2_gap28_pair4_L	TTTCCTCTCAACTAGCCATCG	EF1_2.9_gp8_R_sm	TTCGAAGTAACAAATAAGTCCTCA
EF1_ph2_t6_gap6_p4_L	GAAGGCACTGGCAATAGAGG	EF1_2.9_gp20_L_sm	TATATTCACCGACTAAGGCATTCT
EF1_ph2_t6_gap6_p1_R	CTCAGGGACTCCTGCTTCTG	EF1_2.9_uc102_B_sm	AGATGAGTAATAACGCTAAAACAG
EF1_ph2_t6_gap7_p8_L	GCAGATACGAATGAAAGTTGAGAG	EF1_2.10_gp20_L_sm	TCGAGCGAGTCAATATCCGTAAA
EF1_ph2_t6_gap7_p3_R	TGTTGTAACCTGGAAGGGG	EF1_2.10_uc102_B_sm	ACAAGATAGTGGCGATAGCAAAGA
EF1_ph2_t6_gap13_p2_L	ATTTGTTTTGAGGTGGCGAG	EF1_2.10_uc102_A_sm	TTCCACATATGCAACAAACTCG
EF1_ph2_sm_gap17_L	ACGGTCTCTCGCTATGTCAGTATCG	EF1_2.10_gp17_L_sm	TCAGGGCAAATTTAGTAAGGTGT
EF1_ph2_sm_gap17_R	ATCGTATTCGGAAGTACTGCT	EF1_2.10_gp17_R_sm	AGTTGGCTGAGATGGTTGAGAG
EF1_ph2_t6_gap21_p1_L	AGGTCAACTCGGTGTCCAAG	EF1_2.10_gp8_L_sm	TGACACGTGGCAATTCTACCTTA
EF1_ph2_t6_gap21_p3_R	TTAGCCCCTCATCATTCAGC	EF1_2.10_uc094_B_sm	CTCACCAAAACGAATGTAAAGAA
EF1_ph2_t6_gap22_p3_L	GAATCACGTGCATTTTCATGG	EF1_2.10_gp8_R_sm	ACAACAATATCAGCGCTCAATCCA
EF1_ph2_t6_gap27_p10_L	GACAAGCCTTGCAAAGCAAC	EF1_2.11_gp17_L_sm	AGCTACTTCATGTGTCAACC
EF1_ph2_t6_gap27_p5_R	GTGCGAATACTGTTATCGGC	EF1_2.11_gp17_R_sm	GAACAAGAAGACGTGAAAGAGCA
EF1_ph2.1_sm_gap9_L	TGAGGCTTGTGAGGCGAGTTCTAC	EF1_2.11_uc091_A_sm	TGCCTCCTGTGATGTTTTCTGTGG

EF1_ph2.1_sm_gap9_R	GCACAGGCAAGACTCCATATAGCA	EF1_2.11_gp8_L_sm	CGCCACGATTGCATTATTTGTAT
EF1_ph2.1_sm_gap13_R	CAAAAGATCCAACGTTAGAGAAGC	EF1_2.12_gp17_L_sm	AAATTTTGCCACTAGCGGGGTTCT
EF1_ph2.1_sm_gap14_L	CGAGTCCGTCCTAACAGTCATTTG	EF1_2.12_gp17_R_sm	ACAGGGCAAGAAACGGAAGTG
EF1_ph2.1_sm_gap14_R	TGTCTGTCGCCTGGTTGTAGCACT	EF1_2.13_gp17_L_sm	ACATCCGTTCCCACTATCAAAT
EF1_ph2.2_gap3_L_sm	CGAACCAAATGTGTAGGGCGTCTT	EF1_2.13_gp17_R_sm	TCCCTTACGATGCCGAGACACT
EF1_ph2.2_gap3_R_sm	TCCTTCTCGCTATCGCCACCAC	EF1_2.13_gp6_R_sm	ACGATCGCTTCTAATGGGTTGTCT
EF1_ph2.2_gap25_L_sm	TAAGTCCATCCCGGTCTCTCGTA	EF1_2.13_gp8_L_sm	AACTTCTTGTGCCACTTCTAAT
EF1_ph2.2_gap25_R_sm	CGAGCGCAGGCGGTTTCTTA	EF1_2.14_uc094_B_sm	TTCCGTATGCTATTCTGATTGTCC
EF1_ph2.2_gap26_L_sm	ATCATCTTCTTTTCGCATTTTG	EF1_2.14_uc113_B_sm	ACCACTCGTTTTAGATTGCTTTAC
EF1_ph2.2_gap26_R_sm	TTATCCATTACAGAAGCCCACGAA	EF1_2.14_gp6_R_sm	CGCATCGATTACTTGTCTCTTC
EF1_ph2.3_gap6_L_sm	TATAATGGATCTGAGGGGAACTT	EF1_2.14_uc123_A_sm	ACTCATGAACGGCTTGTAAT
EF1_ph2.3_gap6_R_sm	TGTACCAACCGGGAGCGTCAATA		

13.3 Primer für das Fosmid-Screening

EfFosSc_p2_Cont010203_End_F	GCATAACGAAGCAGGAGGAG
EfFosSc_p2_Cont010203_End_R	TTGTTCTTTGACCGCATCTG
EfFosSc_p2_Cont040506_Beg_F	ATTTGGAAGAGCCTGGACCT
EfFosSc_p2_Cont040506_Beg_R	CAAACCATACTTTTCTTTTATGTGC
EfFosSc_p2_Cont040506_End_F	TGGCTTCCATGCCTTTAATC
EfFosSc_p2_Cont040506_End_R	TGCAGATGGTCTCACTCTGG
EfFosSc_p2_Cont07_Beg_F	AATGCGATGGCCAAAGTAAC
EfFosSc_p2_Cont07_Beg_R	GTATCGGCGCTGGTCTAATC
EfFosSc_p2_Cont27_End_F	GACAAGCCTTGCAAAGCAAC
EfFosSc_p2_Cont27_End_R	CGGGCCTTAGATTCAATCAG
EfFosSc_p2_Cont28_Beg_F	ATTGTTTGGGGCTGAAAATG
EfFosSc_p2_Cont28_Beg_R	GGTGCATGTTGCAGAACAAT
Ef_FS_Cont08_End_F	ATTCGCCATGTTAAGCCACT
Ef_FS_Cont08_End_R	AATGGGGGATAAGACGGTTC
Ef_FS_Cont09_Beg_F	GGACCAGATCCAGACACGTT
Ef_FS_Cont09_Beg_R	TGGAATGTAATCGGCTCTCC
Ef_FS_uc102A_F	GGCCTGGTATTCTTCCCAAT
Ef_FS_uc102A_R	CCAGATGGCGATCGAGTAAT
Ef_FS_Cont2122232425_Beg_F	TCCATCATGGTCGTTTAGCA
Ef_FS_Cont2122232425_Beg_R	GTTGTTTCCGTGTCAACGTG
Ef_FS_uc091_A_F	CCATTGGGACAGGCTTATGT
Ef_FS_uc091_A_R	GTCAGCAAAACACCGGAAAT
Ef_FS_Cont08_Beg_F	TTGTGTTTCCACCATTTGGAT
Ef_FS_Cont08_Beg_R	GAATGGGTGGTATTGGGTTG
Ef_FS_Cont20_End_F	GGGGAAGAATTCATGGCTAA
Ef_FS_Cont20_End_R	CCGAACCAAAGGCTAAATGA

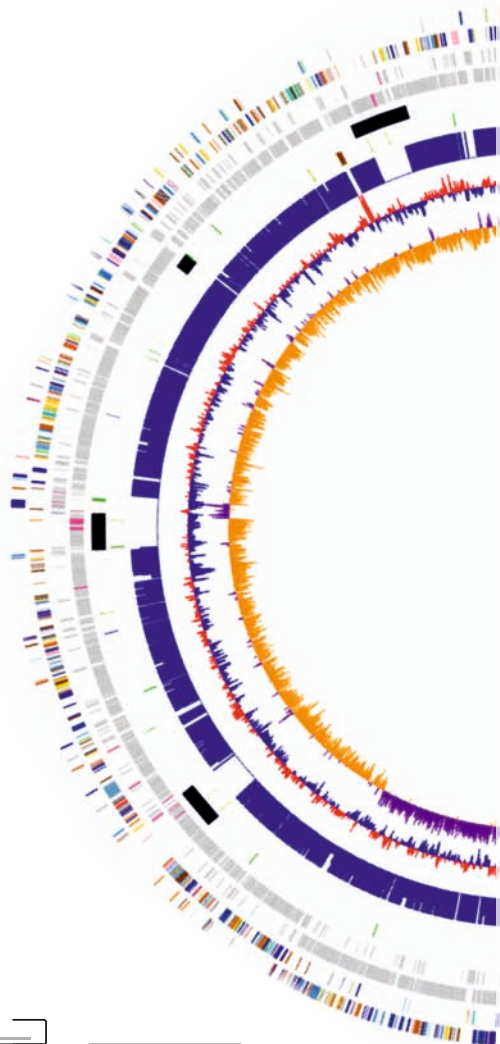
13.4 Primer für die rRNA-Kassetten

16_f_27	GAGTTTGATCCTGGCTCA	23_f_1	GGTTAAGTTAGAAAGGGCGCA
16_f_41	GCTCAGGACGAACGCTGGCG	23_f_2	ACGGCGGAACACGTGAAATTC
16_f_358	AGACTCCTACGGGAGGCAGCAGT	23_f_3	ACTTGGAGATAGCTGGTTCTC
16_r_336	ACTGCTGCCTCCCGTAGGAGTCT	23_f_4	GTCAGACCGGAAGGACTGGTG
16_f_536	GTGCCAGCAGCCGCGGTAATAC	23_f_5	GAGAATCCTAAGGTGAGCGAG
16_r_515	GTATTACCGCGGCTGCTGGCAC	23_f_6	GTGGAGCTTTACTGCAACCTG
16_f_928	TGAAACTCAAAGGAATTGACGGGG	23_f_7	AGGTTTGGCACCTCGATGTC
16_r_907	GCCCCGTC AATTCTTTGAGTTT	23_f_8	GAAAGTAAGATCCCTGAAAG
16_r_1066	ACATCTCACGACACGAGCTG	23_r_1	AACGCTTACGTACAGCTCCC
16_f_new1	GGAGGAAGGTGGGGATGACG	23_r_2	GAATTTACGTGTTCCGCCGT
16_r_1385	CGGTGTGTACAAGGCCC	23_r_3	GAGAACCAGCTATCTCCAAGT
16_f_1385	GGGCCTTGACACACCG	23_r_4	CACCAGTCCTTCGGTCTGAC
		23_r_5	CTCGCTCACCTTAGGATTCTC
		23_r_6	CAGGTTGCAGTAAAGCTCCAC
		23_r_7	GACATCGAGGTGCCAAACCT
		23_r_8	TGGTTAAGTCCTCGATCGAT

13.5 Enterokokken des DMSZ

Type Strains des DMSZ					
Name	Referenz	DSM Nr.	Synonyme	Isoliert aus	Risikogruppe
<i>E. aquimarinus</i>	Svec et al. 2005	17690		Seewasser	1
<i>E. asini</i>	de Vaux et al. 1998	11492		Zökum des Esels	1
<i>E. avium</i>	(ex Nowlan and Deibel 1967) Collins et al. 1984	20679	" <i>Streptococcus avium</i> "	menschl. Fäzes	2
<i>E. caccae</i>	Carvalho et al. 2006	19114		menschl. Fäzes	1
<i>E. canintestini</i>	Naser et al. 2005	21207		Fäzes des Hundes	1
<i>E. canis</i>	De Graef et al. 2003	17029		Hund, Anal-Abstrich	1
<i>E. casseliflavus</i>	(ex Vaughn et al. 1979) Collins et al. 1984	20680	<i>Streptococcus casseliflavus</i> Vaughan et al. 1979, <i>Enterococcus flavescens</i> Pompei et al. 1992	Pflanzenmaterial	1
<i>E. cecorum</i>	(Devriese et al. 1983) Williams et al. 1989	20682	<i>Streptococcus cecorum</i> Devriese et al. 1983	Zökum des Huhns	1
<i>E. columbae</i>	Devriese et al. 1993	7374		Darm der Taube	1
<i>E. devriesei</i>	Svec et al. 2005	22802		bovin	2
<i>E. dispar</i>	Collins et al. 1991	6630		klinisch-menschliche Quelle	2
<i>E. durans</i>	(ex Sherman and Wing 1937) Collins et al. 1984	20633	<i>Streptococcus durans</i> Knight et al. 1984	getrocknete Milch	2
<i>E. faecalis</i>	(Andrewes and Horder 1906) Schleifer and Kilpper-Bälz 1984	20478	<i>Streptococcus faecalis</i> Andrewes and Horder 1906		2
<i>E. faecium</i>	(Orla-Jensen 1919) Schleifer and Kilpper-Bälz 1984	20477	<i>Streptococcus faecium</i> Orla-Jensen 1919		2
<i>E. gallinarum</i>	(Bridge and Sneath 1982) Collins et al. 1984	20628	<i>Streptococcus gallinarum</i> Bridge and Sneath 1982	Darm des Huhns	2
<i>E. gilvus</i>	Tyrrell et al. 2002	15689		menschliche Galle, cholezystitis	1
<i>E. haemoperoxidus</i>	Svec et al. 2001	15920		Betriebswasser	1
<i>E. hermanniensis</i>	Koort et al. 2004	17122		Mandeln des Hundes	1
<i>E. hirae</i>	Farrow and Collins 1985	20160			2
<i>E. inusitatus</i>	?	?	?	?	?
<i>E. italicus</i>	Fortina et al. 2004	15952	<i>Enterococcus saccharominimus</i> Vancanneyt et al. 2004	Toma Käse	1
<i>E. malodoratus</i>	(ex Pette 1955) Collins et al. 1984	20681		Gouda Käse	1
<i>E. moraviensis</i>	Svec et al. 2001	15919		Quelle in Hajek (Publikation: Oberflächenwasser)	1
<i>E. mundtii</i>	Collins et al. 1986	4838		soil	1
<i>E. pallens</i>	Tyrrell et al. 2002	15690		human peritoneal dialysate	1
Type Strains des DMSZ		DSM Nr.			Risikogruppe
Name	Referenz		Synonyme	Isoliert aus	
<i>E. raffinosus</i>	Collins et al. 1989	5633	.	blood culture	2

<i>E. ratti</i>	Teixeira et al. 2001	15687	.	neonatal rat with diarrhea	2
<i>E. saccharolyticus</i>	(Farrow et al. 1985) Rodrigues and Collins 1991	20726	<i>Streptococcus saccharolyticus</i> Farrow et al. 1985	straw bedding	1
<i>Lactococcus garvieae</i>	(Collins et al. 1984) Schleifer et al. 1986	20684	<i>Streptococcus garvieae</i> Collins et al. 1984, <i>Enterococcus seriolicida</i> Kusuda et al. 1991	bovine mastitis	2
<i>E. silesiacus</i>	Svec et al. 2006	22801		drinking water (surface water)	1
<i>Tetragenococcus solitarius</i>	(Collins et al. 1989) Ennahar and Cai 2005	5634	<i>Enterococcus solitarius</i> Collins et al. 1989	ear exudate	2
<i>E. sulfureus</i>	Martinez-Murcia and Collins 1991	6905		plant material	1
<i>E. termitis</i>	Svec et al. 2006	22803		gut of a termite	1
<i>E. thailandicus</i>	Tanasupawat et al. 2008	21767		fermented sausage (mum)	1
<i>E. villorum</i>	Vancanneyt et al. 2001	15688	<i>Enterococcus porcinus</i> Teixeira et al. 2001	piglet, 2-4 days old	2



édition scientifique
VVB LAUFERSWEILER VERLAG

VVB LAUFERSWEILER VERLAG
STAUFENBERGRING 15
D-35396 GIESSEN

Tel: 0641-5599888 Fax: -5599890
redaktion@doktorverlag.de
www.doktorverlag.de

ISBN: 978-3-8359-6051-0



9 78 3 8359 6051 0

